1.c.
20.9.71
[illegible]
26/[illegible]/71

PSYCHOLOGICAL REVIEW

GEORGE MANDLER, EDITOR
JEAN M. MANDLER, ASSOCIATE EDITOR
UNIVERSITY OF CALIFORNIA, SAN DIEGO

Vol. 78, No. 1 **January 1971**

Articles

Editorial . . . . **George Mandler** 1

The "Superstition" Experiment: A Reexamination of Its Implications for the Principles of Adaptive Behavior
J. E. R. Staddon and Virginia L. Simmelhag 3

The Drive Theory of Social Facilitation
Robert Frank Weiss and Franklin G. Miller 44

Punishment: Method and Theory . . . . . **Philip J. Dunham** 58

On the Exponents in Stevens' Law and the Constant in Ekman's Law . . . . **Robert Teghtsoonian** 71

Letters to the Editor 81

Published Bimonthly
by The American Psychological Association, Inc.

Editorial Consultants

Anne Anastasi
Norman H. Anderson
A. John Arrowood
Albert Bandura
Gerald S. Blum
Robert C. Bolles
Gordon H. Bower
Dana Bramel
J. Merrill Carlsmith
A. Charles Catania
Clyde H. Coombs
T. N. Cornsweet
Nicholas B. Cottrell
J. A. Deutsch
Bruce Earhard
M. Earhard
Howard Egeth
Seymour Epstein
E. J. Fantino
Wendell R. Garner
Rochel Gelman
Michael T. Ghiselin
James J. Gibson
Irving L. Janis
Frank A. Logan
Elijah Lovejoy
Norman Miller
Bennett B. Murdock, Jr.
Paul H. Mussen
M. Frank Norman
E. C. Poulton
Karl H. Pribram
Irvin Rock
Eliot H. Rodnick
Richard Rudner
Paul F. Secord
Roger N. Shepard
Norman J. Slamecka
Richard L. Solomon
S. S. Stevens
John Theios
Michael A. Wallach
Robert B. Zajonc

(A large group of psychologists has kindly agreed to assist in the editing of manuscripts submitted to the *Psychological Review*. In each issue the Editor will list in the *Review* those editorial consultants who have helped him process manuscripts for that issue, so that their efforts may be gratefully acknowledged.)

The *Psychological Review* is devoted to articles of theoretical significance to any area of scientific endeavor in psychology. Except for occasional articles solicited by the Editor, manuscripts exceeding 7,500 words will not be accepted. Ordinarily, manuscripts which consist primarily of original reports of research or of reviews of the literature should be submitted to other journals. Methodological and statistical articles are not appropriate for the *Psychological Review*. See the editorials in the *Review* (1965, **72**, 1–2; 1971, **78**, 1–2) for further details.

Beginning with the 1971 issue, Letters to the Editor will be published on an experimental basis. Letters not exceeding 250 words on topics of theoretical interest should be submitted in duplicate.

Manuscripts must be accompanied by an abstract of 100 to 120 words typed on a separate sheet of paper. The abstract should conform to the style of *Psychological Abstracts*. Detailed instructions for preparation of the abstracts appeared in the *American Psychologist* (1961, **16**, 833), or may be obtained from the APA Central Office.

All manuscripts must be submitted in triplicate. In the reference list, give journal titles in full; do not abbreviate. Figures should be submitted as glossy prints; duplicate figures may be photographic or pencil-drawn copies. Authors should retain a copy of the manuscript in case there is loss in the mail.

Because of the large number of manuscripts submitted, there is an inevitable publication lag of several months. Authors may avoid this delay if they are prepared to pay the costs of publishing their own articles; the appearance of articles by other contributors is not thereby delayed.

Manuscripts should be addressed to the Editor, George Mandler, Department of Psychology, University of California, San Diego, La Jolla, California 92037.

Copyrighted material. The following policies govern reprinting or other forms of reproduction of materials copyrighted by the American Psychological Association: Written permission must be obtained for copying or reprinting any tables, figures, or text of more than 500 words in length. Persons or organizations who obtain permission to reproduce any such copyrighted materials must include on the first page of the reproduced materials the exact copyright notice which appeared on the Association's original publication. There will be a charge of $10.00 per page for such permission except when the requester is the author of the article to be reprinted, or when the article is to be reproduced in a limited number of copies solely for instructional purposes.

Helen Orr
Managing Editor

Lorraine Biros
Technical Editor

Henry S. Salmon
Advertising and Business Manager

Robert D. Gorman
Subscription Manager

Published Bimonthly by the
AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
Prince and Lemon Streets, Lancaster, Pa.
1200 Seventeenth St., N. W., Washington, D. C. 20036

$10.00 volume

$2.00 issue

Second-class postage paid at Lancaster, Pa.

Send all communications, including address changes, to 1200 Seventeenth St., N. W., Washington, D. C. 20036. Address changes must arrive by the tenth of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

PSYCHOLOGICAL
REVIEW

GOVT OF W. BENGAL
LIBRARY
CALCUTTA
BUREAU OF EDU
406
4. 11. 71
Lib

VOL. 78, NO. 1 JANUARY 1971

EDITORIAL

A rather tenuous tradition opens the pages of psychological journals to incoming editors. At times these editorials are used to signal the high hopes of the virgin editor, and at others to announce changes in policy, format, or procedures. I am taking advantage of the opportunity offered an incoming editor with some trepidation.

Our editorial system and its editorial and publication lag permits the incoming editor to write this editorial at a time when he has already spent more than six months processing manuscripts. The editorial mantle was transferred from Charles N. Cofer, completing six highly successful and notable years as editor of the *Psychological Review*, to the writer on February 1, 1970. In looking over the editorial written by my predecessor and published in the January 1965 issue of the *Psychological Review*, I find there adequate information on a variety of topics that I would not wish to change.

The editorial policies announced by Dr. Cofer in 1965 seem to be still applicable to the *Psychological Review* in 1971. The prospective contributor is therefore directed to that editorial on such issues as the inappropriateness to the *Psychological Review* of literature reviews, statistical articles, methodological articles, and empirical papers. The guidelines stated there on those issues will continue to be used.

At the same time, I would like to make some additional comments, which reflect no real change in policy but a possible specification of editorial attitudes, as well as some additions and innovations in terms of content and format of the *Psychological Review*.

The *Psychological Review* continues to be the APA journal concerned with psychological theory and its development. As far as theory is concerned, the *Review* will primarily be concerned with the encouragement of theories that are unequivocal, determinate, and wherever possible, more than speculative. Psychological theories should, at their best, also reflect the cumulative nature of the psychological enterprise. At the same time, different fields vary in the degree to which they are able to generate, at this point in time, logically or mathematically unequivocal theories. However, as a goal, such lack of equivocation must always be in the mind of both the author and the editor. In addition, and with Dr. Cofer's initial comments, a manuscript submitted to the *Psychological Review* should reflect a detailed working out of some theory or implication thereof. Minor comments, which may reflect a critical statement or an elaboration, will be acceptable for the Theoretical Notes which will be continued, and prospective contributors might keep in mind the distinction between a regular contribution and a theoretical note.

The development of psychological theory very frequently impinges on historical concerns. The *Psychological Review* will consider for possible publication manuscripts that deal with historical themes, but only if such historical analysis is primarily concerned with theoretical issues. In addition, articles should make some potential contribution to ongoing psychological theory and research. The criterion of impact on current theory and research is also one that will be applied to articles dealing with the philosophy of science as it impinges on psychology. Philosophical disquisitions which display neither

© 1971 by the American Psychological Association, Inc.

Bureau of Edni. & Psyl.
(C. E. R. T.)
Acc. J. 841

intended nor actual implications for current or future psychological theory and research will be discouraged. On the other hand, theoretical articles dealing with practical issues, whether they be social, interpersonal, or technological, will be encouraged. There is no reason to believe that psychological theory need be alienated from the issues facing our society.

There will be three innovations introduced on an experimental basis in 1971. The first is a column entitled The State of the Art. In The State of the Art we shall invite commentary on the current state of psychological theory. Those invited will be, on the one hand, limited in number, since we will, at most, have one such contribution per issue, and they will be leaders in theoretical thought and empirical research in a broad spectrum of psychological activities. This column is designed in the first instance to provide published commentary on current psychological theory and, second, to generate discussion and dialogue among psychologists. As a second step in that direction, we have introduced a Letters to the Editor section. Letters to the Editor in the *Psychological Review* are not intended to be archival publications. Their format will be letter formats, and their titles typically will be selected by the Editor in order to provide information rather than references. Within the body of a letter, again in order to avoid citable or archival qualities, references may be made to published material as the author sees fit, but never as a separate list of references at the end of the letter. It is hoped that these letters will provide brief comments on issues relating to psychological theory, commenting on published articles, or on thoughts and ideas generated by the State of the Art column. Typically, letters to the Editor will not be refereed, but decisions will be made directly by the Editor, with the hope that publication lag for the Letters column can be kept to less than six months. Both for letters and for theoretical notes, a scheme originally developed by E. G. Boring in *Contemporary Psychology* will be applicable, namely, that a reply to an article, note, or letter should not occupy more than 50% of the space of the item to which it constitutes a reply. Such a method will, on the one hand, make discussions possible, but prevent them from becoming interminable.

Finally, the Editor reserves the right to publish editorials in issues other than the first and the last of his editorship. It is my hope that in these editorials some reflections on the editorial process, the state of psychological theory as reflected in the submissions to the Journal, and other issues may be discussed. The Editorial is one of the few rewards of the editorial position, but I do not expect to take advantage of it too often.

Finally, it should be noted that the quality of a journal is a direct reflection of the quality of psychological thought going on in the field. It does not require any humility to say that editors do not significantly direct the nature or quality of thinking in their fields. We shall thus rely on the quality of the contributions and on the extensive help from referees who will, as in the past, be the best available critical reviewers in any field or area in psychology. And it is obviously to these reviewers that both my hopes and, already, my thanks go at this time.

All of these suggestions, thoughts, and innovations are highly tentative. If any of them work to the benefit of the psychological community, they will be continued. If they do not, they will quickly be abandoned. In the meantime, I can only look forward to the ather unhappy prospect of losing friends, making enemies, and battling editorial lags, he common fate of all editors.

GEORGE MANDLER

Psychological Review
1971, Vol. 78, No. 1, 3-43

THE "SUPERSTITION" EXPERIMENT:

A REEXAMINATION OF ITS IMPLICATIONS FOR THE PRINCIPLES OF ADAPTIVE BEHAVIOR[1]

J. E. R. STADDON[2] AND VIRGINIA L. SIMMELHAG[3]
Duke University *Scarborough College, University of Toronto*

Replication and extension of Skinner's "superstition" experiment showed the development of two kinds of behavior at asymptote: *interim activities* (related to adjunctive behavior) occurred just after food delivery; the *terminal response* (a discriminated operant) occurred toward the end of the interval and continued until food delivery. These data suggest a view of operant conditioning (the terminal response) in terms of two sets of principles: principles of behavioral variation that describe the origins of behavior "appropriate" to a situation, in advance of reinforcement; and principles of reinforcement that describe the selective elimination of behavior so produced. This approach was supported by (*a*) an account of the parallels between the Law of Effect and evolution by means of natural selection, (*b*) its ability to shed light on persistent problems in learning (e.g., continuity vs. noncontinuity, variability associated with extinction, the relationship between classical and instrumental conditioning, the controversy between behaviorist and cognitive approaches to learning), and (*c*) its ability to deal with a number of recent anomalies in the learning literature ("instinctive drift," auto-shaping, and auto-maintenance). The interim activities were interpreted in terms of interactions among motivational systems, and this view was supported by a review of the literature on adjunctive behavior and by comparison with similar phenomena in ethology (displacement, redirection, and "vacuum" activities). The proposed theoretical scheme represents a shift away from hypothetical "laws of learning" toward an interpretation of behavioral change in terms of interaction and competition among tendencies to action according to principles evolved in phylogeny.

The field of learning has undergone increasing fractionation in recent years. Interest in "miniature systems" and exact theories of local effects has grown to the detriment of any attempt at overall integration. Consequently, as one perceptive observer has noted:

> At times, one senses a widespread feeling of discouragement about the prospects of ever getting clear on the fundamentals of conditioning. Attempts to arrive at firm decisions about alternative formulations rarely produce incisive results. Every finding seems capable of many explanations. Issues become old, shopworn, and disappear without a proper burial [Jenkins, 1970, pp. 107–108].

The present article outlines an attempt to redress this imbalance. It is organized around the problem of "superstitious" behavior, originated by Skinner some years ago, which plays a crucial part in the empirical and theoretical foundations of current views of learning. Discussion of a replication of Skinner's original experiment leads to an account of the relationships between evolution and learning, and a system of classification derived therefrom. The paper concludes with a theoretical account of "superstition" and some related phenomena.

---

[1] Research was supported by grants from the National Institute of Mental Health, United States Public Health Service, and the National Science Foundation to Duke University and the University of Toronto. The authors thank Nancy K. Innis for advice and assistance. Janice Frank, Irving Diamond, Carl Erickson, Richard Gilbert, and Peter Klopfer kindly commented on earlier versions of this paper. A shorter version of the experimental part of this work was presented in partial fulfillment of the requirements for the MA degree at the University of Toronto (Simmelhag, 1968).

[2] Requests for reprints should be sent to J. E. R. Staddon, Department of Psychology, Duke University, Durham, North Carolina 27706.

[3] Now at Duke University.

THE "SUPERSTITION" EXPERIMENT

In his classic experiment on "superstitious" behavior, Skinner (1948) showed that the mere delivery of food to a hungry animal is sufficient to produce operant conditioning. The pigeons in that experiment were allowed 5-second access to food every 15 seconds, and food delivery was independent of their behavior. Nevertheless, nearly every pigeon developed a recognizable form of stereotyped, superstitious behavior that became temporally correlated with food delivery as training progressed.

Skinner's (1961) analysis of this phenomenon is a straightforward application of the Law of Effect:

> The conditioning process is usually obvious. The bird happens to be executing some response as the hopper appears; as a result it tends to repeat this response. If the interval before the next presentation is not so great that extinction takes place, a second "contingency" is probable. This strengthens the response still further and subsequent reinforcement becomes more probable [p. 405].

Skinner's observations were quickly repeated in a number of laboratories. The apparent simplicity and reliability of the phenomenon, coupled with the plausibility of Skinner's interpretation of it, and the more exciting attractions of work on reinforcement schedules then developing, effectively stifled further study of this situation. However, both the experiment and his explication played a crucial role in advancing Skinner's theoretical view of operant beavior as the strengthening of unpredictably enerated ("emitted") behavior by the autoatic action of reinforcers.

Two kinds of data obtained in recent years aise new questions about "superstition" this sense. First, experiments with timeelated reinforcement schedules have shown e development of so-called "mediating" ehavior during the waiting period, when e animal is not making the reinforced reonse. Thus, on schedules which require e animal to space his responses a few secds apart if they are to be effective in procing reinforcement (spaced-responding edules), pigeons often show activities ch as pacing and turning circles. Similarly, on fixed-interval schedules, in which the first response t seconds after the preceding reinforcement is effective in producing reinforcement, pigeons may show a similar behavior during the postreinforcement "pause" when they are not making the reinforced response. Other species show activities of this sort in the presence of appropriate environmental stimuli; for example, schedule-induced polydipsia, in which rats reinforced with food on temporal reinforcement schedules show excessive drinking if water is continuously available (Falk, 1969). None of these activities is reinforced, in the sense of being contiguous with food delivery, yet they are reliably produced in situations similar in many respects to Skinner's superstition procedure. Possibly, therefore, some of the activities labeled superstitious by Skinner, and attributed by him to accidental reinforcement of spontaneously occurring behavior, may instead reflect the same causal factors as these mediating activities.

Second, a number of experiments have demonstrated the development of behavior in operant conditioning situations by a process more reminiscent of Pavlovian (classical) conditioning than Law of Effect learning as commonly understood. Breland and Breland (1961) reported a series of observations showing that with continued operant training, species-specific behavior will often emerge to disrupt an apparently well-learned operant response. In the cases they describe, behavior closely linked to food (presumably reflecting an instinctive mechanism) began to occur in advance of food delivery, in the presence of previously neutral stimuli ("instinctive drift"). Since these "irrelevant" activities interfered with food delivery by delaying the occurrence of the reinforced response, they cannot be explained by the Law of Effect. A description in terms of stimulus substitution—a principle usually associated with Pavlovian conditioning—is better, although still not completely satisfactory. More recently, Brown and Jenkins (1968) have shown that hungry pigeons can be trained to peck a lighted response key simply by illuminating the key for a few seconds before food delivery. Fewer than a hundred light–food pairings are usually sufficient to bring about key pecking. The relationship between this "auto-shaping" procedure and Pavlovian conditioning is further emphasized by an experiment reported by Williams and Williams (1969). They found that auto-shaped key pecking is maintained even if the key peck turns off the light on the key and thus prevents food delivery on that occasion. All these experiments show the occurrence of food-related behaviors, in anticipation of food, under conditions more or less incompatible with the Law of Effect.

The auto-shaping procedure is operationally identical to Pavlovian conditioning with short delay (the light–food interval in these experiments is typically 8 seconds). Therefore the eventual emergence of food-related behavior, in anticipation of food delivery, is not altogether surprising—although the directed nature of key pecking has no counterpart in principles of conditioning that take salivation as a model response. The superstition situation is also equivalent to a Pavlovian procedure—in this case temporal conditioning, in which the UCS (food) is simply presented at regular intervals. Perhaps, therefore, prolonged exposure to this situation will also lead to the emergence of food-related behavior in anticipation of food. Possibly the superstitious behavior described by Skinner includes activities of this sort, that occur in anticipation of food, as well as mediating activities that occur just after food delivery.

The present experiment affords an opportunity to test these ideas. It provides comparative data on the effect of fixed versus variable interfood intervals on superstitious responding (Skinner used only fixed intervals), as well as allowing a comparison between response-dependent and response-independent fixed-interval schedules. The experiment also extends Skinner's work by recording in some detail both the kind and time of occurrence of superstitious activities. The emphasis is on the steady-state adaptation to the procedures, but some data on the course of development of superstition are presented.

We hope to show that careful study of the superstition situation makes necessary a revision of Skinner's original interpretation and, by extension, requires a shift of emphasis in our view of adaptive behavior.

Method

Subjects

Six pigeons were used: four white Carneaux, two with experimental experience (Birds 31 and 29) and two experimentally naive (Birds 47 and 49). Two other pigeons were of a local (Toronto, Ontario) breed and were experimentally naive (Birds 40 and 91). All the birds were maintained at 80% of their free-feeding weights throughout.

Apparatus

Two standard Grason-Stadler operant conditioning chambers were used. The response keys were covered with white cardboard except during the response-dependent condition when one key was exposed and transilluminated with white light. Data were recorded by a clock, digital counters, and an event recorder. Food delivery was controlled automatically by relays and timers. Behaviors were recorded via push buttons operated by an observer. A tape recorder was used to record comments and corrections. Except for the push buttons and the tape recorder, all programming and recording apparatus was located in a separate room. White noise was present in the experimental chamber and, together with the noise of the ventilating fan, served to mask extraneous sounds.

Procedure

Three schedules of food delivery were used: (*a*) A response-independent fixed-interval (FI) schedule in which the food magazine was presented at 12-second intervals. (*b*) A response-independent variable-interval (VI) schedule in which the food magazine was presented on the average every 8 seconds. The following sequence of interreinforcement intervals was used, programmed by a loop of 16-millimeter film with holes punched at appropriate intervals: 3, 6, 6, 12, 9, 7, 3, 10, 21, 6, 5, 11, 8, 5, 3, 9, 7, 9, 5, 13, 3, 8, 9, 4, 7, 12, 11, 3, 6, 5, and 9 seconds. (*c*) A response-dependent FI schedule in which food was delivered (reinforcement occurred) for the first key peck 12 seconds or more after the preceding reinforcement.

Food delivery involved 2-second access to mixed grain. Sessions ended after the sixty-fourth food delivery and the pigeons were run daily.

Habituation sessions. All the birds were given a number of daily 10-minute sessions when no food was delivered; the birds were simply placed in the chamber and their behavior observed and recorded. The birds received the following numbers of such sessions: Bird 31, 3; Bird 29, 3; Bird 47, 15; Bird 49, 15; Bird 40, 7; Bird 91, 7. Note that the experimentally naive birds received more habituation exposure.

Response-independent training. Four of the pigeons were then given a number of sessions on each of the two response-independent procedures, either FI followed by VI, or the reverse. The birds received the procedures in the indicated order (number of sessions in parentheses): Bird 31: FI 12 (26), VI 8 (111); Bird 29: VI 8 (26), FI 12 (109); Bird 47: VI 8 (36), FI 12 (36); Bird 49: FI 12 (37), VI 8 (36).

Response-dependent training. Two of the birds already trained on the response-independent procedures were then switched to the response-dependent FI 12 for the following numbers of sessions: Bird 31, 37; Bird 47, 52. The two naive birds (40 and 91), after their habituation sessions, were given one session of response-independent FI 12 when the food magazine operated every 12 seconds, but contained no grain. This was to habituate these birds to the sound of the mechanism. The next day these two birds were placed in the experimental box with the food magazine continuously available for 10 minutes. The following day the two birds were introduced to the response-dependent FI 12-second schedule. A small piece of black tape was placed on the lighted response key, as an inducement to pecking, for this first session only. All the birds pecked the key during the first session of the response-dependent procedure. This rather elaborate key-training procedure was designed both to prevent the experimenter from shaping the naive birds' behavior, and to avoid the possibility of "superstitious" conditioning, which might be entailed by some form of auto-shaping. Bird 40 received a total of 45 sessions of the response-dependent FI 12-second schedule, Bird 91 received 38.

Response description and scoring. Response categories were arrived at on the basis of initial observation during the habituation sessions and were altered as necessary to accommodate new behaviors. The names, descriptions, and numbers of the categories appear in Table 1. Responses were scored (by pushing the appropriate button) in two ways, either discretely or continuously. If a response tended to occur in discrete units (e.g., pecking), then the appropriate button was pushed each time an instance of the response occurred. The observer was the same throughout (VLS), and the maximum recordable rate for discrete responses was 3–4 per second. A continuous response is one which took an indefinite amount of time (e.g., facing magazine wall); the appropriate button was pressed throughout the duration of a continuous response. Discrete responses were pecking (wall, key, or floor) and quarter circles, all the rest were continuous responses. In general, the response categories were mutually exclusive. The only exception is facing Magazine wall (R_1), which at various times occurred with Flapping wings (R_5), Moving along magazine wall (R_8), and Pecking (R_7).

Results

The data of interest in this experiment are the kind and amount of behavior at different points in time following the delivery of food. As training progressed, a systematic pattern of behavior as a function of postfood time began to emerge. The properties of this steady-state pattern for VI and FI schedules is discussed first, followed by a description of the changes that took place during acquisition.

Steady-State Behavior

In the steady state, the behavior developed under both the FI and VI procedures fell reliably into two classes: (*a*) The *terminal response* was the behavior that consistently occurred just before food delivery. It began 6–8 seconds after food delivery on the FI procedures, and about 2 seconds after food on the VI procedure, and usually continued until food delivery. (*b*) A number of activities usually preceded the terminal response in the interval. These activities are probably indistinguishable from what has been termed mediating behavior, but we prefer the more descriptive term *interim activities.* These activities were rarely contiguous with food.

Figure 1 shows the performance averaged across three sessions of steady-state responding under all conditions for all the pigeons. The left-hand panels show the response-dependent FI schedule; the middle panels, the response-independent (superstitious) FI schedule; and the right-hand panels, the response-independent VI procedure, for the four birds exposed to each. The graphs show the probability (relative frequency) with which each of the activities occurred during each second of postfood time. Each bird shows the clear division between terminal and interim activities already alluded to. Excluding R_1, and the results for Bird 29 on VI, Pecking (R_7) was the terminal response for all the response-independent procedures. For Bird 29, the terminal response Head in magazine (R_{11}) became an interim activity following the shift from VI to FI and was replaced as terminal response by Pecking. Pecking remained the terminal response for this bird throughout a sequence

TABLE 1

DESCRIPTION OF OBSERVED ACTIVITIES

Response no.	Name	Description
R_1	Magazine wall	An orientation response in which the bird's head and body are directed toward the wall containing the magazine.
R_2	Pecking key	Pecking movements directed at the key.
R_3	Pecking floor	Pecking movements directed at the floor.
R_4	¼ circle	A response in which a count of one ¼ circle would be given for turning 90° away from facing the magazine wall, a count of two for turning 180° away, three for 270°, and four for 360°.
R_5	Flapping wings	A vigorous up and down movement of the bird's wings.
R_6	Window wall	An orientation response in which the bird's head and body are directed toward the door of the experimental chamber containing the observation window.
R_7	Pecking	Pecking movements directed toward some point on the magazine wall. This point generally varied between birds and sometimes within the same bird at different times.
R_8	Moving along magazine wall	A side-stepping motion with breastbone close to the magazine wall, a few steps to the left followed by a few steps to the right, etc. Sometimes accompanied by (*a*) beak pointed up to ceiling, (*b*) hopping, (*c*) flapping wings.
R_9	Preening	Any movement in which the beak comes into contact with the feathers on the bird's body.
R_{10}	Beak to ceiling	The bird moves around the chamber in no particular direction with its beak directed upward touching the ceiling.
R_{11}	Head in magazine	A response in which at least the beak or more of the bird's head is inserted into the magazine opening.
R_{12}	Head movements along magazine wall	The bird faces the magazine wall and moves its head from left to right and/or up and down.
R_{13}	Dizzy motion	A response peculiar to Bird 49 in which the head vibrates rapidly from side to side. It was apparently related to, and alternated with, Pecking (R_7).
R_{14}	Pecking window wall	Pecking movements directed at the door with the observation window in it.
R_{15}	Head to magazine	The bird turns its head toward the magazine.
R_{16}	Locomotion	The bird walks about in no particular direction.

of response-independent procedures after the ones reported here, lasting for a total in excess of 90 sessions. A curious idiosyncratic head movement accompanied pecking by Bird 49 (R_{13}: Dizzy motion) on response-independent FI, although it disappeared following the switch to VI. The locus of Pecking on the magazine wall differed from bird to bird and varied both across and within sessions for some birds. The stable features of this response were its topography and its restriction to the general area of the magazine wall. A variety of interim activities occupied the early parts of the FIs and the period within 2 or 3 seconds after food on the VI schedule: Pecking floor (R_3), ¼ circles (R_4), Flapping wings (R_5), Moving along magazine wall (R_8), and Beak to ceiling (R_{10}) were the most frequent. The interim activities were therefore more variable from bird to bird than was the terminal response.

The pattern of behavior characteristic of each interval was little affected by whether or not food was dependent on key pecking. The similarity between the patterns during response-dependent and response-independent FI is particularly striking for Bird 47, who was switched directly from the response-independent to the response-depend-

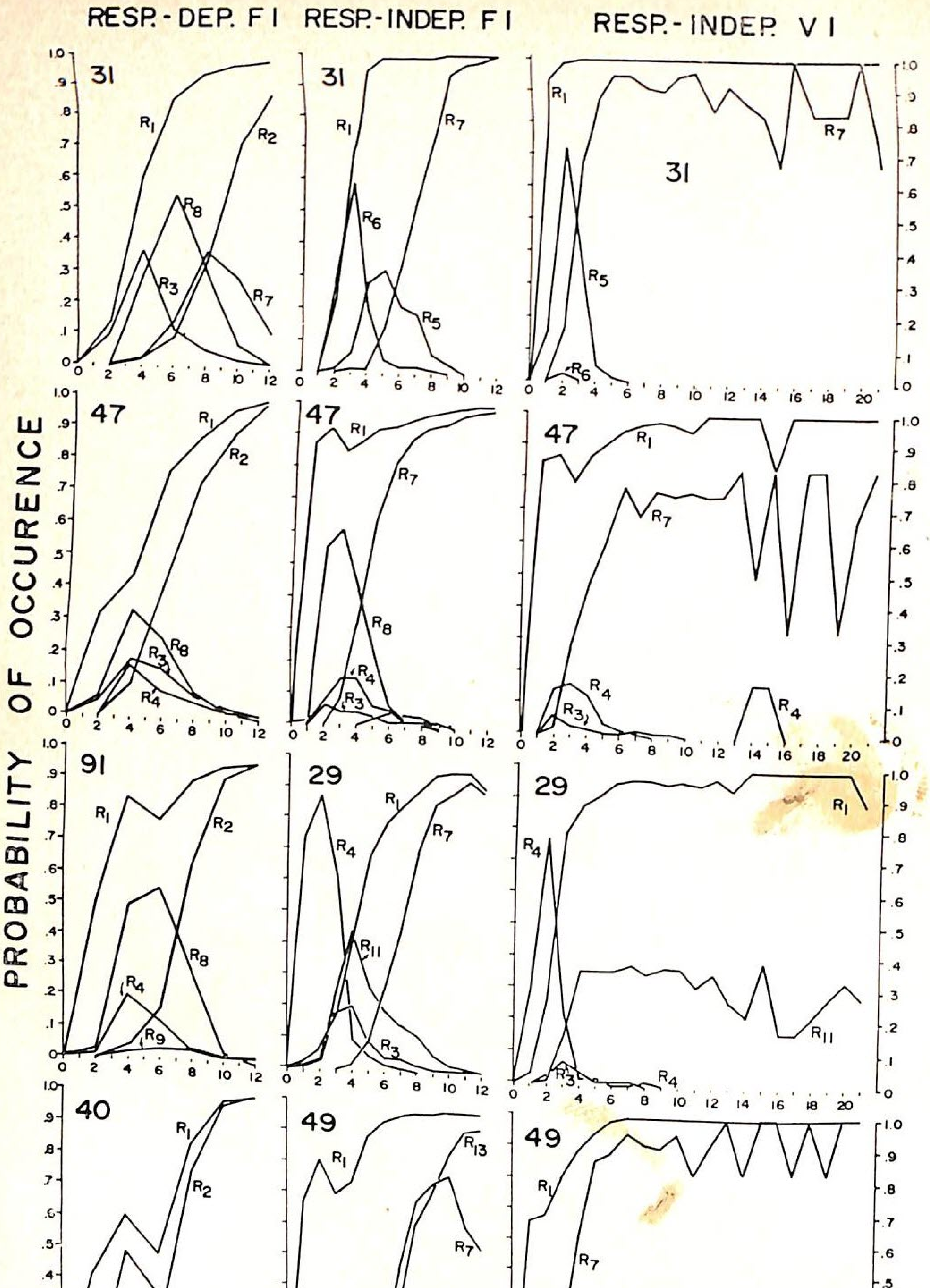


Fig. 1. Probability of each behavior as a function of postfood time for all birds for all three experimental conditions, averaged over three sessions of steady-state responding under each condition. (Each point gives the probability that a given behavior occurred in that second of postfood time. Data for the response-dependent condition are averaged across 2-second blocks. Behaviors (R_i) are identified in Table 1.)

nt procedure. Bird 31, for whom the reponse-independent and response-dependent 'Is were separated by response-independent 'I, shows more variation in the interim ctivities under the two conditions. Birds 1 and 40, who were exposed only to the esponse-dependent FI, show a similar patern of interim activities to the responsendependent birds, although again there is ome variation as to details. This similarity annot be attributed to the procedure used shape key pecking (see Procedure). After ne imposition of the key-pecking continency, Bird 31 retained the old Pecking reponse as an interim activity restricted to a eriod in the interval just before key peckng. Bird 47 showed a similar effect of his esponse-independent experience in that a igh proportion of his key pecks failed to epress the key sufficiently to activate the utomatic recording circuitry, although they ere recorded as key pecks by the observer. verall, these data provide no evidence for ubstantial changes in the pattern of terminal interim activities traceable to the imposion of a key-pecking requirement.

The general pattern of terminal and inrim activities during the response-indeendent VI schedule was similar to the FI ocedures. Differences were restriction of e interim activities to the first 2 or 3 secds of postfood time rather than to the first or 7 seconds (with one exception, to be dissed), and the smaller number of interim tivities, in most cases. Under both fixed d variable procedures, once the terminal sponse began in an interval, it continued til food delivery. The exception is Bird who showed a drop in the probability of e terminal response (Pecking, R_7) accomnied by a transient increase in the interim tivity of $\frac{1}{4}$ circles (R_4) at the 14–15-cond postfood time. A similar slight drop the probability of Pecking, although not companied by an interim activity, was also own by Bird 31. These differences are reed to the properties of the VI schedule ee Discussion).

quential Structure

Figure 2 indicates something of the seential structure of the behavior occupying each interfood interval for each bird during the response-independent procedures. The figure summarizes the two to five behavior sequences that account for most of the intervals during the three steady-state sessions. For simplicity, no account is taken either of interbehavior times or of the duration of each activity in this method of representation. The most striking characteristics of these sequences are: (*a*) that each bird showed only a small number of typical sequences (usually three or four); (*b*) that the sequencing was very rigid, so that although a given behavior might fail to occur during a particular interval, it never occurred out of sequence—this is indicated by the absence of return arrows ("loops") in these diagrams; and (*c*) that the variability of the sequences was greatest early in the interval and least at the end, in the period just preceding food delivery—this is indicated by the absence of "forks" (ambiguous transitions from one behavior to two or more others, as in the diagram for Bird 31 where $R_6 \rightarrow R_5$ or $R_6 \rightarrow R_7$ with approximately equal probability) late in the sequence of behaviors shown in the diagrams. This regular sequencing did not occur early in training, as indicated in Figure 4, discussed below.

An inviting possibility raised by these regular sequences is that this behavior may be described by some kind of Markov chain (cf. Cane, 1961). Although the argument cannot be presented in full here, this assumption cannot be sustained for a number of reasons, the most important of which are (*a*) that the duration of a bout of a given activity was shorter the later the activity began within an interval, and (*b*) that the time between two successive activities was shorter the later in the interval the first activity ended. These and other considerations suggest that postfood time was the most important factor controlling both the onset and offset of each activity in the sequence.

Acquisition

Figure 3 shows acquisition data for naive Bird 49, through and beyond the period when his behavior became stable on re-

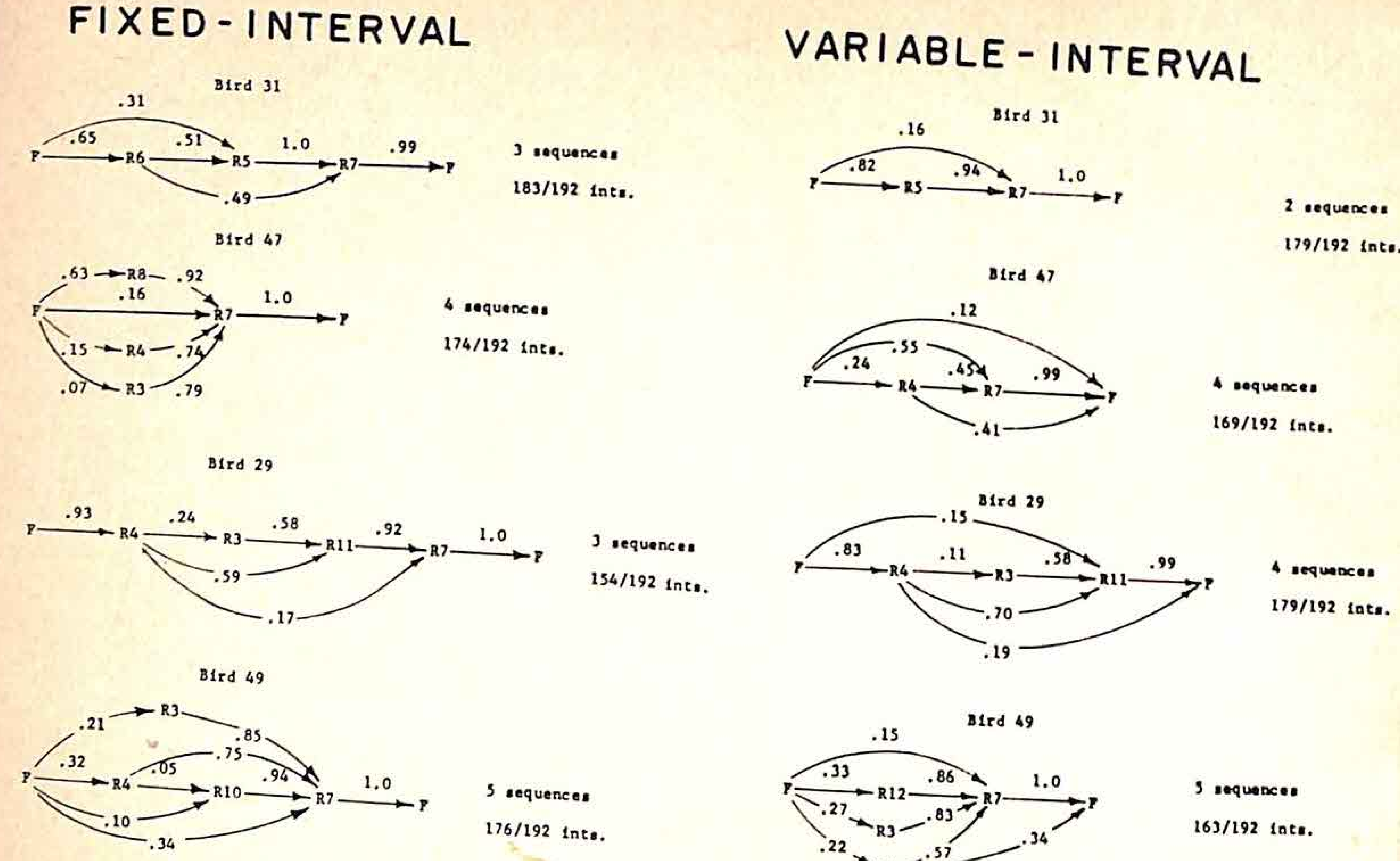


FIG. 2. Steady-state sequences: Sequential relationships among behaviors during the last three sessions of each response-independent procedure for all four pigeons. (Fixed-interval is on the left, variable-interval on the right. Numbers give probabilities of the indicated transitions. Number of different sequences on which the diagrams are based and number of intervals (out of 192) accounted for by these are indicated. "F" is food delivery, and behaviors (R_i) are identified in Table 1. Note: probabilities at each "fork" do not always sum to 1.0 because not every sequence is accounted for.)

sponse-independent FI. The graphs show (as a function of sessions) the probability of the various behaviors in each of the six 2-second periods making up the FI. The most noteworthy characteristic of acquisition is the relatively sudden disappearance of the behavior Head in magazine, which was almost the only behavior to occur during the first few sessions for Bird 49, in favor of the terminal response Pecking on magazine wall. For Bird 49 this transition took place between the seventh and eighth sessions of FI without any prior history of pecking during earlier sessions. Once pecking became established as a terminal response, the only further change was a slow decline in the probability of the response during early parts of the interval (e.g., between 3 and 8 seconds). The other birds showed similar results on their first exposure to the response-independent procedures; for each bird, pecking on the magazine wall first occurred during the following sessions: for Bird 31, on the first FI session; Bird 29, twelfth session of FI following the switch from response-independent VI; Bird 47, twenty-eighth session of VI; Bird 49, eighth session of FI. Thus, although one of the experienced birds pecked during the first response-independent session (31), the other did not until being switched to a different schedule (29). No bird shifted to a different terminal response once he began pecking; the total number of sessions of response-independent experience (FI and VI) for each of the four birds following the onset of pecking was as follows: Bird 31, 140; Bird 29, 96; Bird 47, 45; Bird 49, 66. For Bird 29, the terminal response during the response-independent FI procedure (i.e., before the onset of pecking) was Head in magazine. As with the other birds, once pecking appeared it was in full strength almost immediately and further experience

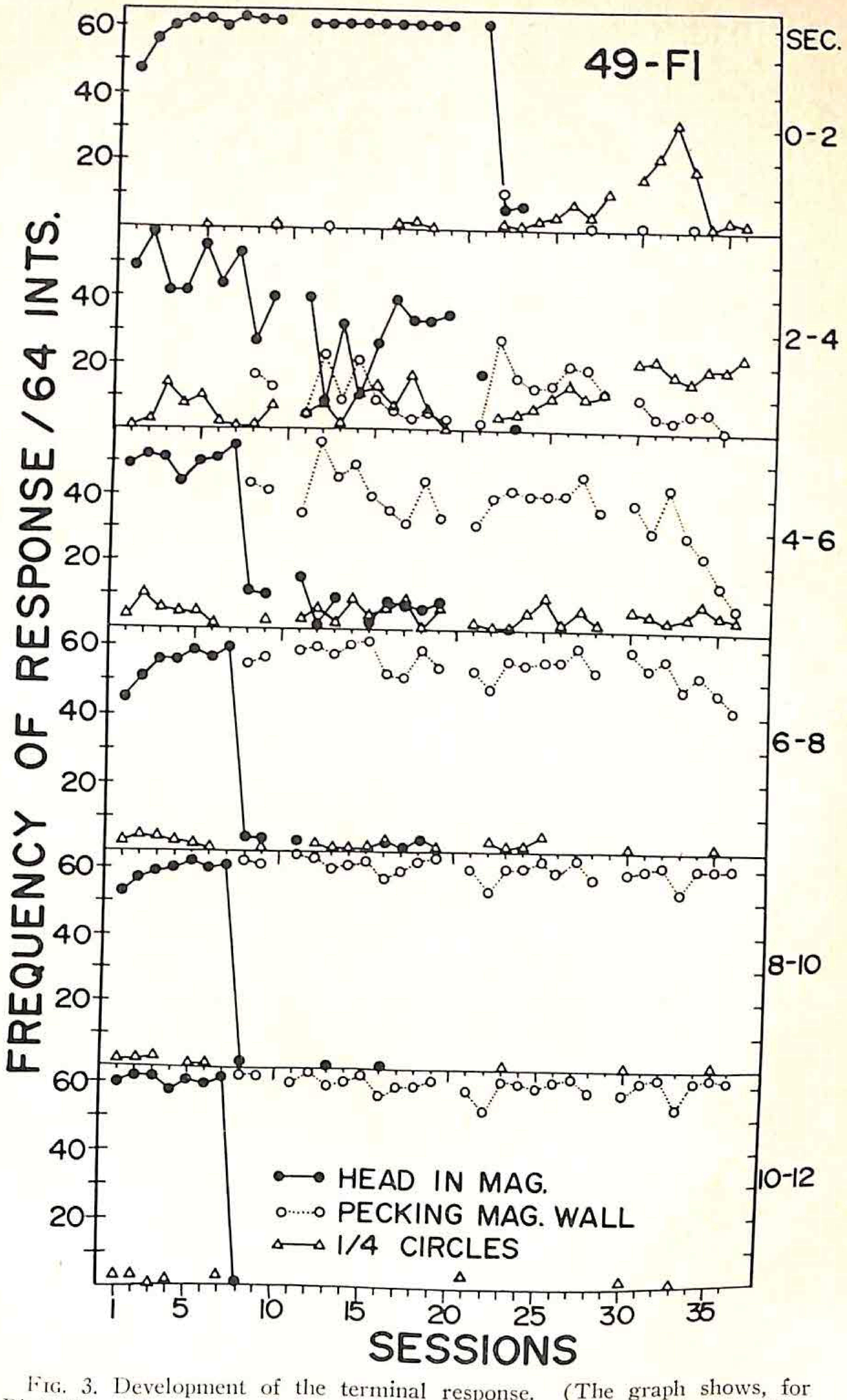


FIG. 3. Development of the terminal response. (The graph shows, for Bird 49 on the response-independent fixed-interval procedure, the transition from Head in magazine (R_{11}) to Pecking (R_7) as terminal response, and includes one interim activity, ¼ circles (R_4), for comparison. Each panel covers 2 seconds of the 12-second interval and indicates the number of intervals (out of 64) in which a given response occurred in that 2-second block for each session over the first 36 sessions. Gaps indicate days for which data are not available. Bird 49 was not run for a 9-day period between Sessions 21 and 22.)

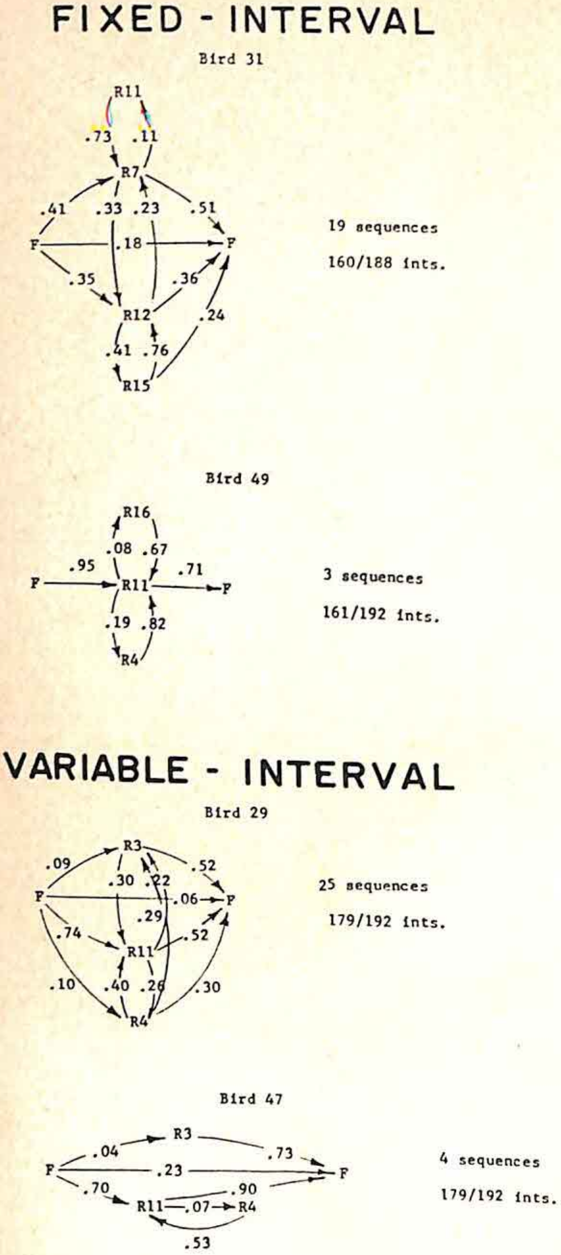


Fig. 4. Sequence data, in the same form as Figure 2, for the first three sessions under the first response-independent procedure to which each bird was exposed. (Details as in Figure 2.)

had the effect merely of restricting it more to the later parts of the interval.

Figure 4 shows sequence data, in the same form as Figure 2, for the first three sessions of the first response-independent schedule to which each bird was exposed. Approximately the same proportion of the total number of sequences observed within 192 intervals is accounted for by these diagrams as in the diagrams of Figure 2 for the last three sessions. The data for the first three sessions are much more variable, showing repetitions of a behavior within an interval and reversals of sequence; three out of four birds show no single terminal response. Thus, the regularities apparent in Figure 2 are evidently a real effect of training and not simply an artifact of this method of representation.

Discussion

The results of this experiment confirm the suggestion that the "superstition" situation generally produces two distinct kinds of activity: interim activities that occur at short and intermediate postfood times, and the terminal response that begins later in the interval and continues until food delivery. It is not always clear from Skinner's original discussion just which kind of behavior he was observing. In one case, he briefly describes an experiment in which the interfood interval was 60 seconds, and the "superstitious" response (a "well-defined hopping step from the right to the left foot") was automatically recorded. In this case, the behavior was evidently a terminal response, since it occurred with increasing frequency through the interval:

> The bird does not respond immediately after eating, but when 10 or 15 or even 20 sec. have elapsed it begins to respond rapidly and continues until the reinforcement is received [Skinner, 1961, p. 406].

On other occasions, however, Skinner may have been observing interim activities, as, in our experience, they are sometimes much more striking than the terminal response, especially early in training when they may include actions like jumping in the air, vigorous wing flapping, and idiosyncratic head and limb movements. We have also sometimes observed interim activities during sessions when there was no obvious terminal response.

Nature of the Terminal Response

The data from both FI and VI schedules of food presentation indicate that the probability of the terminal response at different postfood times was a function of the probability of food delivery at those times: the probability of the terminal response increased just in advance of increases in the probability of food delivery, and decreased just after decreases in that probability. Thus, on the FI schedule, the terminal response began at postfood times greater than about 8 seconds, corresponding to a probability of food delivery of zero at times less than 12 seconds, and one thereafter. On the VI schedule, probability of food delivery was zero at postfood times less than 3 seconds (the shortest interfood interval in the VI sequence), and increased thereafter. Correspondingly, the birds began their terminal response after 2 seconds or so of postfood time and continued until food delivery. The data for Birds 29 and 47 on the VI schedule provide additional confirming evidence. Both birds showed a small decline in the probability of the terminal response between Seconds 12 and 17 of postfood time, accompanied by a brief reappearance of an interim activity (between Seconds 13 and 16) in the case of Bird 47. This brief decrease in the terminal response, and accompanying reappearance of an interim activity, corresponds to the zero probability of food in the region of postfood time between 13 and 21 seconds (because the VI sequence used had no interfood interval between 13 and 21 seconds in length—see Method). Catania and Reynolds (1968) showed a similar relationship between rate of key pecking and probability of reinforcement, as a function of postreinforcement time, on a variety of conventional (response-dependent) VI reinforcement schedules. This similarity and the similarities between the response-independent and response-dependent conditions of this experiment emphasize that the terminal response must be regarded as a discriminated operant in Skinner's (1958) sense.

Herrnstein (1966) reported an experiment in which the rate of key pecking on response-independent FI 11 seconds was lower than on response-dependent FI 11. However, in the present experiment, considering just the terminal response of Pecking on magazine wall (R_7), we found no evidence for a difference in either probability or rate of response favoring the response-dependent procedure. Indeed, rate of key pecking (defined as switch operations) under the response-dependent FI was generally lower than the observer-defined rate of pecking under the response-independent FI because many pecks failed to break the switch contact. On the other hand, because the location of pecking was much more variable under the response-independent condition, a comparison between key pecking (in the response-dependent condition) with pecking on a comparable area of the magazine wall (in the response-independent condition) would show more responding under the response-dependent condition. This result raises the possibility that the effect of the response-dependency, in operant conditioning experiments using interval reinforcement schedules, may be largely one of determining (perhaps imperfectly) the *location* of pecking, rather than either its form or its frequency of occurrence.

Nature of the Interim Activities

Falk (1969, 1970) has coined the term *adjunctive behavior* for a variety of activities that are induced in a number of different animal species (rats, pigeons, monkeys, chimpanzees) by intermittent schedules of reinforcement. These activities generally occur just after reinforcement, when the reinforced (i.e., terminal) response is not occurring. However, the important variable in determining their temporal location seems to be the low probability of further reinforcement in the immediate postreinforcement period (rather than the interruption of eating), since they also occur following brief stimulus presentations on second-order interval schedules (Rosenblith, 1970), following each response on spaced-responding (differential reinforcement of low rates) schedules (Segal & Holloway, 1963), during timeout periods (Wüttke, 1970), only after the last pellet when a number are delivered consecutively at the end of each FI (Keehn, 1970), and later in the interval during long FI schedules (Segal, Oden, & Deadwyler, 1965). They are not elicited (e.g., by frustration) in any obvious sense, since they take some time to develop (Reynierse &

Spanier, 1968). They are related to motivational systems, since a number of activities—for example, polydipsia (excessive drinking), pica (eating nonfood material), wheel running, and schedule-induced aggression—may occur more or less interchangeably depending on the presence of appropriate stimuli; and animals will learn an operant response in order to obtain an opportunity to engage in one of these activities (Falk, 1970).

The interim activities in the present experiment appear to reflect the same causal factors as adjunctive behavior: they occur at times when reinforcement is not available and the terminal response is not occurring; they occur on both response-dependent and response-independent schedules (cf. Azrin, Hutchinson, & Hake, 1966; Burks, 1970; Flory, 1969); and they occur on both FI and VI schedules. Adjunctive behavior requires appropriate stimuli (water for polydipsia, wood shavings for pica, etc.) to which it is directed and by which it can be modified, within limits. This is unlikely to be a crucial difference, however, since appropriate stimuli were not available in the present experiment (had water been available, schedule-induced drinking would almost certainly have been observed in lieu of the interim activities; cf. Shanab & Peterson, 1969), and behavior resembling interim activities has occasionally been reported under conditions when most animals show adjunctive behavior, thus

> [this rat] was atypical in that it did not develop polydipsia . . . exhibiting instead a number of stereotyped behaviors, like rearing and running to the corners of the experimental chamber, between reinforcements [Keehn, 1970, pp. 164–167].

We return to a theoretical account of interim and adjunctive behavior in the concluding section.

An analogy can be drawn between the terminal and interim activities here and the classical dichotomy between consummatory and appetitive behavior (Craig, 1918). Thus, pecking, the stable terminal response, is a food-elicited (consummatory) activity in pigeons, and the interim activities were quite variable, as might be expected of appetitive behavior. Moreover, the variability of the sequences (as measured by the number of "forks" in the sequence) was greatest at the beginning of the sequence and decreased toward the end (cf. Figure 2). More less unlearned sequences terminating in c summatory acts show a similar reduct in variability toward the end of the seque (e.g., Morris, 1958).

In a similar vein, Falk (1970) has com pared adjunctive behavior with displacement activities (Tinbergen, 1952):

> In both adjunctive behavior and displacement activity situations, the interruption of a consummatory behavior in an intensely motivated animal induces the occurrence of another behavior immediately following the interruption [p. 305].

At the present stage of knowledge, these comparisons do little more than group together a number of puzzling phenomena that cannot as yet be convincingly explained, either by ethological principles or by the Law of Effect. We can choose either to accept these behaviors as anomalies within our present conceptual system, hoping that further research will show how to reconcile them, or revise the system in a way that will accommodate them more naturally. The first alternative is becoming increasingly hard to maintain, as it becomes clear that these behaviors are of wide occurrence, and as they continue to resist attempts to explain them in conventional terms. Polydipsia is the most widely studied adjunctive behavior, and Falk (1969) summarizes research results as follows:

> It is not explicable in terms of any altered state of water balance initiated by the experimental conditions. It cannot be attributed to adventitious reinforcing effects. Since it cannot be related to abnormal water losses, chronic internal stimulation arising from unusual states . . . or from injury to the central nervous system, the overdrinking would be classified clinically as primary or psychogenic polydipsia [p. 587].

A revision of the conceptual foundations of operant behavior, which will deal naturally with these behaviors, as well as guide research into more profitable channels, seems called for.

Development of the Terminal Response

The development of the terminal response provides some clues toward an alternative conception. Changes in the terminal response throughout acquisition rule out an unqualified application of the Law of Effect as a description of the process. For example, three of the four birds made the response Head in magazine at a higher frequency and for a larger fraction of the total time than any other response, for the first few sessions of exposure to the superstition procedure (either FI or VI, depending on the bird). This behavior was also the one most often contiguous with the delivery of food. Yet during later sessions, it dropped out abruptly and was replaced by Pecking. Thus the development of Pecking as terminal response resembles the findings of Williams and Williams and of Breland and Breland, much more than it does the strengthening of an emitted response, in Skinner's terms, by the automatic action of food as a reinforcer. In all these cases, the presentation of food at a predictable time resulted, after training, in the regular occurrence of a food-related behavior in anticipation of food delivery, quite independently of the demand characteristics of the situation (i.e., the reinforcement schedule). Given that food is a stimulus that elicits pecking by pigeons, the appearance of pecking, in anticipation of food, in the present experiment is an instance of the principle of stimulus substitution (Hilgard & Marquis, 1940). The stimulus is, of course, a temporal one (postfood time), and the situation is analogous, therefore, to Pavlovian temporal conditioning (Pavlov, 1927), both operationally and in its conformity to the substitution principle. Therefore, it is tempting to attribute the appearance of pecking to a classical conditioning mechanism, and leave it at that.

A number of considerations suggest that this explanation will not suffice, however: (*a*) The behavior of one bird (29) is a partial exception. He showed a terminal response (Head in magazine) different from Pecking, although this response was metastable (Staddon, 1965), in the sense that it was displaced by Pecking following a schedule shift. In addition, Skinner's results and informal observations in a number of laboratories indicate that the superstition procedure may generate behaviors other than pecking that persist for considerable periods of time. Presumably, many of these behaviors are terminal responses, in our sense, and, metastable or not, they cannot be dismissed in favor of a Pavlovian account of all terminal responses in superstition situations. (*b*) The results of Rachlin (1969), who was able to obtain spontaneous key pecking by means of the auto-shaping procedure of Brown and Jenkins, but using electric shock-reduction rather than food as the reinforcer, also complicate the picture, since it is not clear that pecking is elicited by shock, as it is by food. The results of Sidman and Fletcher (1968), who were able to auto shape key pushing in rhesus monkeys, are also not readily explicable by stimulus substitution. (*c*) Williams and Williams (1969) note that the directed nature of the key peck in the auto-shaping situation is not readily accommodated within a Pavlovian framework:

> the directed quality of the induced pecking does not follow naturally from respondent principles (see also Brown and Jenkins, 1968). It is unclear, for example, why pecking would be directed at the key rather than the feeder, or indeed why it would be directed anywhere at all [p. 519].

A similar objection can be raised here, since the terminal response of pecking was always directed at the magazine wall, although it is important to notice that this objection is more damaging to an interpretation in terms of classical conditioning than one couched simply in terms of stimulus substitution. (*d*) Finally, there is the problem of the skeletal nature of pecking. Ever since Skinner's (1938) original suggestion, it has become increasingly common to restrict the concept of classical conditioning to autonomically mediated responses. Some criticisms of this convention are presented below, and Staddon (1970a) has argued against the operant/respondent and emitted/elicited dichotomies. For the moment, let it be said that the objection to the skeletal nature of the response is only a problem for an interpretation of the development of pecking in terms of classical conditioning, as traditionally conceived. It does not conflict with a stimulus substitution interpretation.

In summary, it is clear that while the principle of stimulus substitution describes the development of the terminal response in a majority of cases, and is more successful than an appeal to Pavlovian principles modeled on the salivation reference experiment, it is not adequate as a universal account. We turn now to the possibility of a general scheme to deal with these anomalous facts.

EVOLUTION AND LEARNING

Objectively considered, the subject matter of the psychology of animal learning—the behavior of animals—is a part of biology. This commonality does not extend to terms and concepts, however. Although the ethologists have investigated unlearned behaviors in a variety of species, learning remains almost exclusively the possession of psychologists. Consequently, the theoretical foundations of the study of learning, such as they are, have evolved almost independently of biology. It is now 15 years since Verplanck (1955) wrote: "the structure of the theory of unlearned behavior and that of learned behavior must prove to be similar if not identical [p. 140]," but little progress toward a unified set of concepts has been made. Yet the facts discussed in the previous section seem to demand an interpretation within the context of adaptive behavior as a whole.

The principles of evolution by natural selection provide a unifying framework for biology. Pfaffman (1970) has recently commented:

> I am impressed by the extent to which evolution, the genetic machinery, and biochemistry provide for biologists a common language and unity of theory that overrides the molecular versus organismic debate within biology. In contrast, it is obvious that there is no unified theory of behavior for all students of behavior [p. 438].

These considerations—the commonality of subject matter between biology and animal psychology, the probable common basis of learned and unlearned behavior, the unifying role of evolutionary processes in biology—all suggest the application of evolutionary principles to the psychology of learning in animals. In recent years, several papers have drawn attention to the similarities between evolution and learning (e.g., Breland & Breland, 1966; Broadbent, 1961; Gilbert, 1970; Herrnstein, 1964; Pringle, 1951; Skinner, 1966a, 1966b, 1969), but no version of the evolutionary approach has proved influential as yet. The main reasons for this failure, perhaps, have been the comparative effectiveness of traditional versions of the Law of Effect in dealing with the limited phenomena of laboratory learning experiments and the lack of a substantial body of facts clearly in conflict with accepted theory. To some extent, of course, the second factor reflects the first—although it would be cynical to speculate that the kind of experiments psychologists do are often such as to preclude data that might go beyond current theory. In any event, neither of these reasons now holds true.

The growing number of facts on autoshaping, instinctive drift, adjunctive, and superstitious behaviors are not readily accommodated within traditional views. In addition, the history of the Law of Effect as a principle of *acquisition* (as opposed to the steady state, i.e., asymptote) has not been a distinguished one. Little remains of the impressive edifice erected by Hull and his followers on this base. We are no longer concerned with the production of learning curves, nor with the measurement of habit strength or reaction potential. Hullian theory has proven effective neither in the elucidation of complex cases nor as an aid to the discovery of new phenomena. Indeed, the opposite has generally been true: new learning phenomena such as schedules of reinforcement, learning, and reversal sets, etc. have typically been the result of unaided curiosity rather than the hypothetico-deductive method, as exemplified by Hull and his students. This represents a failure of Hullian theory rather than a general indictment of the hypothetico-deductive method, which has proven every bit as powerful as Hull believed it to be—when used by others (e.g., Darwin, cf. Ghiselin, 1969). A similar, although less sweeping, verdict must be handed down on stochastic learning theory, which represents perhaps the most direct attempt to translate the Law of Effect into a quantitative principle of acquisition. With the exception of predictions about the steady state, such as Estes' ingenious deduction of probability matching (Estes, 1957), the early promise of this approach as a way of understanding the behavior of individual animals has not been fulfilled. It seems fair to say that the rash of theoretical elaboration that has occupied the years since Thorndike first stated the Law of Effect has told us almost nothing more about the moment-by-moment behavior of a single organism in a learning situation. Although the alternative we will offer may be no better than earlier views, it does accommodate the anomalies we have discussed—and the need for *some* alternative can hardly be questioned.

The close analogy between the Law of Effect and evolution suggests an approach that may be a step toward the general framework which the psychology of learning so obviously lacks. At the present stage of knowledge, this approach is simply an analogy, although a compelling one, and cannot yet be called an evolutionary theory of learning. However, it provides a beginning and may lead to a truly evolutionary account which can securely imbed the study of learning in the broader field of biology.

Behavioral Variation and Reinforcement

The Law of Effect, suitably modified to take account of advances since Thorndike's original formulation, can be stated so as to emphasize acquisition, or the steady state, and learning (e.g., S–R bonds), or performance. We will take as a point of departure a neutral version of the law that emphasizes performance in the steady state, as follows: "If, in a given situation, a positive correlation be imposed between some aspect of an animal's behavior and the delivery of reinforcement, that behavior will generally come to predominate in that situation." The term "correlation" is intended to include cases of delay of reinforcement and experiments in which the response acts on the rate of reinforcement directly (e.g., Herrnstein & Hineline, 1966; Keehn, 1970). This formulation does not take account of more complex situations where more than one behavior and more than one correlation are involved (i.e., choice situations), as these complications do not affect the present argument. A discussion of three aspects of this law—(*a*) the initial behavior in the situation before reinforcement is introduced, (*b*) the process whereby this behavior is transformed into the dominant reinforced behavior, and (*c*) reinforcement—follows:

(*a*) The behavior in a situation before the occurrence of reinforcement reflects a number of factors, including past experience in similar situations (transfer), motivation, stimulus factors (e.g., novel or sign stimuli), and others. We propose the label "principles of behavioral variation" for all such factors that originate behavior. These principles are analogous to Darwin's laws of variation, corresponding to the modern laws of heredity and ontogeny that provide the phenotypes on which selection (analogous to the principles of reinforcement, see below) can act. Thus, the term "variation" is intended to denote not mere variability, but the organized production of novelty, in the Darwinian sense.

(*b*) Transition from initial behavior to final behavior is the traditional problem of learning theory and, as we have seen, is essentially unsolved at the level of the individual organism. However, it is at this point that the analogy to the mechanism of natural selection becomes most apparent. Broadbent (1961) makes the parallel quite clear:

> Since individual animals differ, and those with useful characteristics [with respect to a particular niche] survive and pass them on to their children, we can explain the delicate adjustment of each animal's shape to its surroundings without requiring a conscious purpose on the part of the Polar bear to grow a white coat. Equally each individual animal [under the Law of Effect] tries various actions and those become more common which are followed by consummatory acts [i.e., reinforcement] [p. 56].

Thus, the transition from initial to final behavior can be viewed as the outcome of two processes: a process that generates behavior, and a process that selects (i.e., selectively eliminates) from the behavior so produced. Since there is no reason to suppose that the process which generates behavior following the first and subsequent reinforcements is different in kind from the process that generated the initial behavior (although the *effects* will generally be different, see Extinction, below), we can include both under the head of principles of behavioral variation. We propose the label "principles of reinforcement" for the second, selective process.

(*c*) As we have seen, the Darwinian principle of selection is analogous to the process that transforms initial behavior into final behavior—the "principles of reinforcement." The notion of reinforcement (more exactly, the schedule of reinforcement) itself is in fact analogous to an earlier concept, one that preceded evolution by natural selection and can be derived from it: the Law of Conditions of Existence, that is, the fact that organisms are adapted to a particular niche. This is apparent in the form of the statements: "The Polar bear has a white coat *because* it is adaptive in his environment." "The pigeon pecks the key *because* he is reinforced for doing so." It is important to emphasize this distinction between *reinforcement* and *principles of reinforcement*, and the analogous distinction between adaptation to a niche and the process of selection by which adaptation comes about. In the case of adaptation, the process of selection involves differential reproduction, either through absence of a mate, infertility, or death before reproductive maturity. In the case of reinforcement, the distinction is less obvious, since the principles of reinforcement refer to the *laws by means of which* behaviors that fail to yield reinforcement are eliminated, rather than the simple fact that reinforced behaviors generally predominate at the expense of unreinforced behaviors.

Thus, both evolution and learning can be regarded as the outcome of two independent processes: a process of variation that generates either phenotypes, in the case of evolution, or behavior, in the case of learning; and a process of selection that acts within the limits set by the first process. In both cases, the actual outcome of the total process is related to, but not identical with, the material acted upon: phenotypes reproduce more or less successfully, but a gene pool is the outcome of selection; similarly, behaviors are more or less highly correlated with reinforcement, but learning (i.e., an alteration in memory) results.

The three aspects of this process—variation, selection, and adaptation—have received differing emphases at different times, depending on the prevailing state of knowledge. Before Darwin, adaptation, in the form of the Law of Conditions of Existence, was emphasized, since the only explanation for it—the design of the Creator—was not scientifically fruitful. Following Darwin, selection, both natural and artificial, received increased attention, in the absence of firm knowledge of the mechanism of variation (i.e., inheritance). With the advent of Mendelian genetics, variation has been most intensively studied (this is one aspect of the "molecular vs. organismic" debate referred to by Pfaffman).

In terms of this development, the study of learning is at a relatively primitive level, since the Law of Effect, although a great advance over the level of understanding which preceded it, simply represents the identification of environmental events—reinforcers—with respect to which behavior is adaptive. Lacking is a clear understanding of both selection (the principles of reinforcement) and variation (the principles of variation).

Space precludes exhaustive elaboration of all the implications of the classificatory scheme we are suggesting. However, in order to provide some context for our account of the facts already discussed, it seems essential to briefly summarize some possible candidates for principles of variation and reinforcement. It should be obvious that current knowledge does not permit the categories used to be either exhaustive or clear-cut.

Principles of behavioral variation. 1. Transfer processes: One of the main sources of behavior in a new situation is obviously past experience in similar situations. Transfer has been most exhaustively studied under the restricted conditions of verbal learning (e.g., Tulving & Madigan, 1970), but the principles of memory thus derived—proactive and retroactive interference, primacy and recency, retrieval factors, etc.—are presumably of general applicability.

With few exceptions (e.g., Gonzales, Behrend, & Bitterman, 1967), these principles have been little used to interpret animal learning experiments. Other principles of transfer are stimulus and response generalization (induction) and what might be called "compositional transfer," in which several past experiences are combined to generate a novel behavior, as in insight learning and other forms of subjective organization of past input.

2. Stimulus substitution: This principle, which is usually identified with Pavlovian conditioning, has already been discussed as a description of the origin of the terminal response of Pecking. It may also describe the origin of the metastable terminal response Head in magazine, since this response is also elicited by food under these conditions, and the animal which showed this response most persistently (Bird 29) had had considerable experimental experience. The difference in the persistence of Head in magazine in the case of this bird, as compared with others, cannot be explained in this way and might reflect a difference in other transfer processes. The final dominance of Pecking, in every case, may reflect a special susceptibility of consummatory responses to the stimulus substitution principle, or the action of other transfer principles in this particular situation in ways that are presently unclear. The indefinite persistence of key pecking following only three peck-contingent reinforcements, found by Neuringer (1970b), tends to support the simpler conclusion, as does the data of Wolin (1968), who reports a similarity between the topography of operant pecking for food or water reinforcers, and the appropriate unconditioned response. We return later to a general discussion of this principle in relation to these data (pp. 33–34).

3. Preparatory responses: This principle is also frequently associated with classical conditioning, and in that sense is related to, and to some extent overlaps with, the stimulus substitution principle. Thus, some conditioned responses, such as salivation, can be equally well described by either principle. Others, also respondents (such as heart rate, which increases following electric shock, but usually decreases in anticipation of it [Zeaman & Smith, 1965]) may be classified as preparatory responses. Skeletal responses observed in classical conditioning situations are often preparatory in nature (see discussion of classical conditioning, below).

4. Syntactic constraints: There are often sequential constraints among behaviors, so that a given behavior is determined by some property of the sequence of preceding behaviors. Examples are spontaneous position alternation observed in rats and other rodents, stimulus alternation observed in monkeys on learning set problems (e.g., Levine, 1965), and sequential dependencies observed in most species which cause responses to occur in runs rather than alternating randomly (e.g., position habits and other perseverative errors). Human language provides the most developed example of syntactic constraints.

5. Orienting responses: This category includes all those transient behaviors, such as exploration, play, curiosity, etc., that expose the organism to new stimuli and provide the possibility of transfer to future situations.

6. Situation-specific and species-typical responses: Certain situations seem to call forth specific responses, which are often typical of the species rather than the individual and do not seem to depend in any obvious way on any of the other principles of variation such as transfer, etc. Examples are the species-specific defense reactions discussed by Bolles (1970), which occur in fear-producing situations, the tendency to peck bright objects shown by many birds (Breland & Breland, 1966), and the digging shown by small rodents (Fantino & Cole, 1968). Other examples are given by Glickman and Sroges (1966).

Principles of reinforcement. Before the discovery of the mechanism of inheritance, evolution could be explained only in terms of a goal—adaptation—and a means sufficient to reach that goal—variation and selection; the generation-by-generation details of the process were obscure: "Our ignorance of the laws of variation is profound [Darwin, 1951, p. 170]." We are at present equally ignorant of the mechanisms of behavioral variation. Since learning must usually involve constant interplay between variation and reinforcement, we are not yet in a position to suggest anything specific about the moment-by-moment details of the process, either in its variational or selective aspects. However, just as Darwin was able to say something about selection by pointing out the adaptive role of various structures, so it is possible to learn something about the selective role of reinforcement by looking at steady-state adaptations to fixed conditions of reinforcement, that is, reinforcement schedules. On this basis, we suggest the following tentative generalizations about the effects of reinforcement:

1. Reinforcement acts directly only on the terminal response; activities which occur at other times (interim activities, adjunctive behavior, etc.) must be accounted for in other ways, to be discussed later. This assertion is perhaps closer to a definition than an empirical generalization, since it is equivalent to the assertion that the terminal response may, in general, be *identified* as the activity occurring in closest proximity to reinforcement in the steady state. Identification is, of course, no problem in conditioning situations that *enforce* a contingency between some property of behavior and the delivery of reinforcement. However, we will show later that there is no empirical or logical basis for separating situations that do impose a contingency between response and reinforcement from those that do not (see Classical Conditioning, below).

2. Reinforcement acts only to eliminate behaviors that are less directly correlated with reinforcement than others. This generalization, like the first, is also more like a definition, since all that is *observed* (under consistent conditions of reinforcement) is the eventual predominance of one behavior over others—which is consistent with either a suppressive or a strengthening effect. As Skinner (1966a) points out in a summary of the Law of Effect:

> Thorndike was closer to the principle of natural selection than the [usual] statement of his law. He did not need to say that a response which had been followed by a certain kind of consequence was more likely to occur again but simply that it was not less likely. It eventually held the field because responses which failed to have such effects tended, like less favored species, to disappear [p. 13].

Unfortunately, Skinner, and most other behaviorists, elected to follow Thorndike in considering reinforcement to have a positive, strengthening, or "stamping-in" effect. One of the main purposes of the present paper is to suggest that this decision was a mistake and has given rise to a number of problems and controversies that can be avoided if the effects of reinforcement are considered to be purely selective or suppressive.

There are three kinds of argument which support a purely selective role for reinforcement. The first, and most important, is that the overall conceptual scheme which results is simpler and more easily related to biological accounts of behavior than the alternative.

The second point is that situations such as extinction and shaping by successive approximations that might seem to require an active role for reinforcement can be interpreted in a way that does not require anything more than a selective effect. This point is discussed later (see Extinction, below).

The third point is that the superstition and related experiments suggest that the response contingency imposed by most reinforcement schedules is not essential for the production of *some* terminal response, but only for the selection of one response over others, or for directing a response which would probably predominate in any case—as in key pecking by pigeons. Our failure to find a consistent difference in rate of observer-defined pecking between the response-dependent and response-independent conditions of the present superstition experiment supports this view, as does a recent finding that the contingency between electric shock and responding is not necessary to the generation of behavior maintained by intermittent shock delivery to squirrel monkeys (Hutchinson, 1970; Stretch, personal communication, 1970).

The usual interpretation of the fact that there is a terminal response in the superstition situation, despite the absence of response-contingency, is the notion of accidental strengthening of a response by contiguity with the delivery of reinforcement. Here we consider the general implications of adventitious reinforcement as an explanation, and its incompatibility with the notion of reinforcement as selection. Other problems related to adventitious reinforcement are discussed later (see Acquisition, Classical Conditioning).

First, adventitious reinforcement implies failure of constancy, in the sense that the animal is presumed to be unable, because of the stamping-in mechanism of reinforcement, to distinguish between real and accidental correlations between his behavior and the occurrence of reinforcement. This is a strong assumption, in view of the adaptive utility of the constancy process and its ubiquity in perceptual and motor mechanisms. In perception, a similar failure to distinguish changes in sensory input that are produced by our own behavior from changes that are independent of behavior might cause us to perceive the world as rotating every time we turn our head. It is of course true that on the basis of one or a few instances the animal may not be in a position to be certain about the reality of a contingent relationship between his behavior and reinforcement—and this kind of sampling limitation might account for a transient superstitious effect. It is less convincing as an account of a long-term effect.

Second, if reinforcement is considered as purely selective, it *cannot* be invoked as an explanation of behavior when *no* imposed contingency exists between reinforcement and behavior (i.e., in the *absence* of selection). To do otherwise would be like taking a population of white mice, breeding them for 20 generations without further selection for color, and then attributing the resulting white population to the results of "accidental selection." In this case, as in the case of response-independent reinforcement, the outcome reflects a characteristic of the initial population (i.e., the mice gene pool, the nature of the organism), and not a nonexistent selection process.

In short, the notion of adventitious reinforcement is not a tenable one. The extent to which reinforcement can be invoked as an explanation for behavior is directly related both to the degree of *imposed contingency* between response and reinforcement and to the opportunities for that contingency to have some selective effect (i.e., the number of contingent reinforcements). If there is no contingency, or if few contingent reinforcements have occurred, the resulting behavior must owe more to principles of variation than to the selective action of reinforcement. On the other hand, if many contingent reinforcements have been delivered during a protracted period of shaping, the final form of behavior is obviously much less dependent on *particular* principles of variation, and the role of reinforcement (selection) may properly be emphasized.

3. There is considerable evidence for the generality of a principle implicating relative rate or proximity of reinforcement as the fundamental independent variable determining the spatial and temporal location of responding in steady-state conditioning situations. Thus, Herrnstein (1970) has recently reviewed a number of operant conditioning experiments involving differential reinforcement of simultaneous (concurrent schedules) and successive (multiple schedules) choices which support the idea of relative reinforcement rate as the independent variable most directly related to the rate of key pecking in pigeons. Shimp (1969) has presented an analysis which is formally different from Herrnstein's, but which also implicates differences in reinforcement rate as the crucial variable. An extensive series of experiments by Catania and Reynolds (1968) suggests a similar (although less exact) relationship between rate of pecking and relative temporal density of reinforcement on interval reinforcement schedules. Jenkins (1970) summarizes a series of studies with a discrete-trials procedure that led him to suggest relative proximity to reinforcement as an important determiner of the tendency to respond in the presence of a stimulus. Staddon (1970a) has suggested a similar principle to account for positive "goal gradients" that underlie the effects of reinforcement omission on a variety of interval schedules.

4. The concept of reinforcement implies a capacity to be reinforced. The fact that a

Bureau of Ednl. & Psyl. Research
(S. C. E. R. T.)

given stimulus may be reinforcing at one time, but not at another, requires the idea of a *state* corresponding to each class of reinforcers. The independent variable of which the strength of most of these states is a function is deprivation with respect to the appropriate class of reinforcers (food deprivation for the hunger state, water for thirst, etc.). This will not do for most negative reinforcers (e.g., the removal of electric shock), however, since there is no obvious counterpart to deprivation in this case. It is also unlikely that deprivation is the only independent variable sufficient to alter the strength of states associated with positive reinforcement. For example, evidence is discussed later in favor of reciprocal inhibitory interaction between states as a possible factor in polydipsia and other adjunctive behavior. Interactions of this sort may also alter the strength of states for which there is no deprivation requirement, as in audio anaesthesia (Licklider, 1961).

Thus, one may hope for a set of principles of reinforcement that will deal both with the proper classification of states and with the interactions among them.

The theoretical vocabulary of learning is full of terms with an uneasy conceptual status somewhere between explanation, definition, and category label. This terminology, which is not coherent or internally consistent, makes it difficult to approach particular topics with an open mind. It is too easy to dismiss an experimental result as due to adventitious reinforcement or respondent conditioning without, in fact, having any clear understanding of what has been said. Simply defining everything operationally is of little help in this situation, since a set of definitions is not a theory. And a theory, in the sense of a system of concepts that is internally consistent and coherent, is what is required if we are to be sure, in particular cases, whether we really understand a phenomenon—or are merely substituting one mystery for another, with the assistance of an opaque vocabulary.

What we are proposing is too primitive to be called a theory in this sense. However, it does offer a system of classification that makes it difficult to have the illusion of understanding a phenomenon if comprehension is really lacking. In the following section, some implications of this scheme are shown in three major areas: acquisition, extinction, and classical conditioning. This is followed by a brief discussion of possible difficulties of this approach. With the aid of this groundwork, it will then be easier to return to a general account of the superstition and related experiments in the concluding section.

Acquisition

The number of trials necessary for learning is one of those perennial problems that seems to defy resolution. Appeal to data is not conclusive because learning curves are sometimes incremental and sometimes steplike. Even in particular cases, theory is not conclusive either, since with sufficient ingenuity, theoretical accounts of both kinds of curve may be constructed on the basis of either one (or a few) trial learning assumptions or incremental assumptions involving thresholds. We turn now to the possibility that the issue is a consequence of the stamping-in view of reinforcement and becomes less urgent once that view is challenged.

A comment by Skinner (1953) on the necessary and sufficient conditions for the development of superstition provides an illustration:

> In superstitious operant behavior . . . the process of conditioning has miscarried. Conditioning offers tremendous advantages in equipping the organism with behavior which is effective in a novel environment, but there appears to be no way of preventing the acquisition of non-advantageous behavior through accident. Curiously, this difficulty must have increased as the process of conditioning was accelerated in the course of evolution. If, for example, three reinforcements were always required in order to change the probability of a response, superstitious behavior would be unlikely. It is only because organisms have reached the point at which a single contingency makes a substantial change that they are vulnerable to coincidences [pp. 86–87].

Even within the framework of the stamping-in view, it is clear that the truth of this statement depends on a tacit assumption that responses will not generally occur more than once unless followed by reinforcement.

If a given response can be relied on to occur at least 20 times in succession, even without reinforcement, then 3-trial, or even 10-trial learning might well be sufficient to insure its acquisition under the conditions of the superstition experiment. The assumption that responses will occur only once in the absence of reinforcement is a strong assumption about syntactic constraints, in our terminology. Moreover, it is contradicted by the results of the Williams and Williams study (1969), which show indefinite persistence of pecking in the absence of any contiguous relationship between pecks and reinforcement. There is no reason to suppose that a similar persistence is not characteristic of other behaviors (e.g., position habits), although pecking may be more persistent than most. Thus, the finding of superstitious terminal responses, or of indefinite pecking following just three response-contingent reinforcements (Neuringer, 1970b), need imply nothing about the number of trials necessary for learning.

These considerations suggest, as a minimum, the need to take variation into account in discussions of the "speed of conditioning," since rapid acquisition may either reflect an unpersistent response that is really learned rapidly, or a very persistent one that may be learned quite slowly. No inferences about "speed of conditioning" can be drawn solely on the basis of speed of acquisition without information about the frequency and pattern of a given behavior to be expected in a given situation (which may include predictable delivery of reinforcement) *in the absence of contiguity* between that behavior and reinforcement. In practice, since information of the required sort is rarely, if ever, available, it seems wise to defer the issue of speed of learning until behavioral variation has been much more thoroughly studied.[4]

Thus, the moment-by-moment details concerning the effect of reinforcement remain uncertain until much more is known about variation. In the meantime it seems more parsimonious and less likely to lead to fruitless controversies about "speed of conditioning," continuity versus noncontinuity, etc. to assume that the appearance of one behavior, rather than another, at a certain time or place, rather than some other time or place, always requires explanation in terms of principles of variation, with only the *disappearance* of behaviors being attributable to the effects of reinforcement.

This general approach is not novel. It resembles both Harlow's (1959) account of learning-set acquisition in terms of the progressive elimination of error factors, and certain versions of stimulus-sampling theory (Neimark & Estes, 1967). In Harlow's terms, as in ours, one-trial acquisition is a phenomenon that depends on the existence of factors that make the correct behavior much more probable (and persistent) than others (i.e., upon principles of variation). In the learning-set case, these factors are embodied in the prior training procedure, which progressively selects for an initially weak behavior (the "win stay, lose shift" strategy) at the expense of the initially much stronger tendencies to approach particular stimuli. A lengthy process may not be essential, however, for principles of variation involving insight ("compositional transfer," see above) may serve the same function, if they are available to the animal. The important point is the shift of emphasis away from the supposed efficacy of some stamping-in mechanism, the action of which must remain obscure in the absence of knowledge about variation, to the principles of variation that determine the strength of behaviors in advance of contiguity with reinforcement.[5]

Extinction

Extinction is often used as a test for "what is learned" during a training procedure, as in generalization testing (Guttman & Kalish, 1956), and testing for control by temporal factors (Ferster & Skinner, 1957; Staddon, 1970b). Under these conditions, it is assumed that behavior is determined almost entirely by transfer from the base-line condition. Providing the difference between the extinction and training conditions is not too great, either in terms of environmental factors (the stimulus situation is not too different) or temporal factors (the extinction is not prolonged), this assumption can be justified by the reliability and predictability of the behavior usually observed.

---

[4] Problems of this sort are not solved by referring to a hypothetical "operant level" because (*a*) this level is often zero in the absence of a history of reinforcement in the situation; (*b*) it is rarely constant, as the term level implies; and (*c*) the problem of the origin of this level is thereby simply evaded.

[5] Memory has not been separately discussed in this account of acquisition because it is embodied in most of the variational and selective processes we have described. The argument of the present section suggests that a separate account of memory may have to await advances in our knowledge of these processes.

When these conditions are not satisfied or when the training preceding extinction has not been protracted, this reliability is not usually found. On the contrary, extinction under these conditions is usually associated with an increase in the variability of behavior (Antonitis, 1951; Millenson & Hurwitz, 1961). This increase in variability is exactly what would be expected if, as we have suggested, reinforcement has a purely selective effect: in these terms, training involves a progressive reduction in variability under the selective action of reinforcement (centripetal selection, see below), so that absence of reinforcement (extinction) represents a relaxation of selection—with an attendant rise in variability. We turn now to a brief account of the effects of changes in the amount and direction of selection in evolution, which may shed some further light on the properties of behavioral extinction.

Darwin (1896) comments on the effect of domestication as follows:

> From a remote period to the present day, under climates and circumstances as different as it is possible to conceive, organic beings of all kinds, when domesticated or cultivated, have varied. . . . These facts, and innumerable others which could be added, indicate that a change of almost any kind in the conditions of life suffices to cause variability . . . [Vol. 2, p. 243].

Although Darwin sometimes (erroneously) interpreted this observation as reflecting a direct effect of changed conditions on the reproductive system, it can be interpreted in modern terms as due to a relaxation of selection. This is clear from the concept of *centripetal selection* (Haldane, 1959; Mayr, 1963; Simpson, 1953), which refers to the fact that selection under *un*changing conditions, if long continued, acts to weed out extremes, rather than systematically to shift population characteristics in any particular direction:

> When adaptation is keeping up, selection at any one time will be mainly in favor of the existing type. . . . In such cases, the intensity of selection tends to affect not the rate of change but the amount of variation [Simpson, 1953, p. 147].

Thus, a *change* in conditions will generally involve a shift *away* from centripetal selection, with its tendency to reduce variability, and will often lead, therefore, to increased variability. The most obvious example of the effects of relaxation of selection in evolution is degenerating or vestigial structures, that are no longer being selected for:

> It is so commonly true that degenerating structures are highly variable that this may be advanced as an empirical evolutionary generalization [Simpson, 1953, p. 75].

We have already noted that the onset of variability in extinction is often delayed. A similar delay in the effect of changed conditions is often apparent in evolution, Darwin (1896) notes:

> We have good grounds for believing that the influence of changed conditions accumulates, so that no effect is produced on a species until it has been exposed during several generations to continued cultivation or domestication. Universal experience shows us that when new flowers are first introduced into our gardens they do not vary; but ultimately all, with the rarest exceptions, vary to a greater or less extent [Vol. 2, p. 249].

Similar delays have also been reported in experiments on artificial selection (Mayr, 1963). These delays seem to reflect what has been termed "genetic inertia" or "genetic homeostasis" (Mayr, 1963), that is, the tendency for a gene pool which is the result of a long period of consistent selection to resist changes in the direction of selection. A similar mechanism in behavior might account for the dependence of variability in extinction on the duration of the preceding training period, which was referred to earlier: The amount of variability might be expected to be greater and its onset sooner following a brief training period than after one of longer duration. Genetic homeostasis also seems to be involved in the phenomenon of reversion, to be discussed next.

Not all the variation which occurs either in behavioral extinction, or following a change in the conditions of life in evolution, is wholly novel. A relatively common effect, for example, is the reappearance of what Darwin terms "ancestral types," that is, phenotypes which predominated earlier in phylogeny but which have been selected against more recently. This is the phenomenon of reversion which, because of his ignorance concerning heredity, Darwin (1896) found among the most mysterious of evolutionary processes:

> But on the doctrine of reversion . . . the germ [germ plasm] becomes a far more marvellous object, for, besides the visible changes which it undergoes [i.e., phenotypic expressions], we must believe that it is crowded with invisible characters, proper to . . . ancestors separated by hundreds or even thousands of generations from the present time: and these characters, like those written on paper with invisible ink, lie ready to be evolved whenever the organisation is disturbed by certain known or unknown conditions [Vol. 2, pp. 35–36].

Thus, one effect of a relaxation of selection is a more or less transient increase in the relative influence of the distant past at the expense of the immediate past. In behavioral extinction, this should involve the reappearance of old (in the sense of previously extinguished) behavior patterns; that is, transfer from conditions preceding the training condition at the expense of transfer from the training condition.[6] In both cases, evolution and behavior, the effect of the change in conditions may be expected to depend on variables such as the magnitude of the change and the time since the preceding change.

The analogy from Darwin suggests that any considerable change in conditions should increase variability, yet a change in reinforcement schedules that includes an *increase* in reinforcement rate is not usually thought of as producing an increase in variability. This apparent contradiction is resolved by noting that an increase in rate of reinforcement, in addition to changing conditions, also increases the rate of selection (since the analogy assumes reinforcement to have a purely selective effect). Thus, variability may be briefly increased, but since the rapidity of selection is also increased, the net effect may be small. An analogous (but impossible) phenomenon in evolution would be to decrease the time between generations at the same time that conditions are changed. This would speed up the attainment of a new equilibrium and minimize the increase in variability generally associated with changed environment.

The increase in variability due to extinction is most directly put to use in the process of shaping by successive approximations. Frequently, following the first few reinforcements delivered during a "shaping" session, the effect is simply an increase in the range and vigor of behavior. This change can be viewed as being due to the interruption of eating (cf. Mandler, 1964), however, rather than any direct strengthening effect of reinforcement (which we are questioning in any case). In terms of the foregoing analysis, the conditions following the first reinforcement should be optimal for an increase in variability: the change is large (from continuous eating to absence of food) and the training procedure is of short duration (the 3–4-second eating bout), so that time since the preceding change is also short. As food continues to be delivered intermittently, selection occurs and variability decreases.

We have been suggesting a purely selective (rather than strengthening, stamping-in, or energizing) role for reinforcement. The present discussion suggests that such an essentially passive role is compatible with a number of phenomena—extinction, the activating effects of isolated reinforcements—that may appear to demand a more active role for reinforcement. This compatibility was established by drawing attention to similar phenomena in evolution, where the purely selective effect of the conditions of life (analogous to the schedule of reinforcement) is unquestioned. However, no *necessary* identity between the genetic mechanism, which is responsible for the effect of changed conditions on variability in structure, and whatever process is responsible for analogous effects in behavior, is intended. Any process for the production of variation that incorporates some latent memory of past adaptations is likely to show similar effects.

---

[6] Other than clinical accounts of regression, we have been able to find only one published report of this effect—in an account describing shaping porpoises to show novel behaviors (Pryor, Haag, & O'Reilly, 1969). However, we have frequently observed it while shaping pigeons: if a pigeon has been trained in the past to perform a variety of responses, the increase in variability during extinction of the most recently reinforced response generally includes the reappearance of earlier responses.

Classical Conditioning

Our scheme has strong implications for the distinction between classical (Pavlovian, respondent) and instrumental (operant) conditioning, to the extent that the distinction goes beyond procedural differences. Classical conditioning is often thought of as a paradigmatic instance of the primary process of learning: "The [learning] process appears to be based entirely on temporal contiguity and to have classical conditioning as its behavioral prototype [Sheffield, 1965, p. 321]." The salivation "reference experiment" can be interpreted as prototypical in at least two ways that are not always kept separate. The first (which has some similarities to our position) is referred to by Sheffield—the notion that learning depends solely on temporal relationships. Guthrie's aphorism that the animal "learns what he does" is a related idea. It is not easy to find a definitive account of this position, but it may perhaps be summarized by saying that reinforcement or reward is simply necessary to ensure that some behavior occurs in a conditioning situation. Principles involving temporal relationships (contiguity) then ensure that whatever occurs will transfer from one occasion to the next.

The second way in which classical conditioning is discussed as prototypical is in terms of the rule that relates the conditioned and unconditioned responses. Pavlov (1927) emphasized stimulus substitution as the distinctive property of the situation: the response originally elicited only by the UCS is later made to the CS. Subsequently, two kinds of departure from this rule have been pointed out: (*a*) Even in the salivation experiment, there are other readily identifiable components of the conditioned response that do not fit the stimulus substitution rule. These preparatory responses (Zener, 1937) are largely, but not exclusively, skeletal (rather than autonomic). (*b*) Even in the case of salivation and other autonomic responses, the CR is rarely identical to the UCR (i.e., a redintegrative response), so that components of the UCR may be missing from the CR. More serious are differences in direction of change between CR and UCR, which may not even be consistent across individuals, as in heart rate and respiratory conditioning (Martin & Levey, 1969; Upton, 1929; Zeaman & Smith, 1965).

Partly because of problems involving preparatory responses, classical conditioning has increasingly been restricted to autonomically mediated responses. There were two bases for this restriction: the apparent difficulty of conditioning skeletal responses by the operations of classical conditioning and according to the stimulus substitution principle (cf. Skinner, 1938, p. 115), and the supposed impossibility of conditioning autonomic responses via the Law of Effect. This is clea from Kimble's (1961) comment:

> Obviously the common expression, "*the* conditione response," is misleading, and probably in important ways. At the same time it should be recognized that the behavior described by Zener [preparatory responses] was almost certainly instrumentally, rather than classically, conditioned [54].

The force of this argument is lost once th susceptibility of autonomic responses suc as salivation to operant conditioning is demonstrated.

The foregoing facts are sufficient to sho the error of continuing to regard classic conditioning as a unified process, mu less as an explanatory element in accou of operant conditioning. If classical co ditioning is a single process, then it must describable by principles of operation th apply to every instance. As we have se even stimulus substitution, the most gene such principle, fails to apply in every ca

Many, but not all, the anomalous cases are skeletal responses—which suggested that perhaps the notion of a single process could be preserved by restricting the term to autonomic responses. The only independent basis for this is to segregate skeletal and autonomic responses on the grounds that operant conditioning of autonomic responses is impossible. Since this is now known to be false (Miller, 1969), the only remaining basis for excluding skeletal responses from the class of classically conditionable responses is their failure to conform to the principle of stimulus substitution. But, in addition to being tautologous (classical conditioning simply becomes equivalent to learning via stimulus substitution), this fails because many autonomic responses do not conform to this principle and, based on the work of Brown and Jenkins and of Williams and Williams, at least one skeletal response—pecking in pigeons—does obey it (other possibilities are leg flexion and eyeblink). Thus, the class of classically conditionable responses can be defined neither in terms of the neural mediating system (autonomic vs. skeletal) nor in terms of adherence to a particular principle of learning.

The only remaining feature common to all the situations labeled as classical conditioning is the procedure itself. Research in this area has tended to focus on the properties of the temporal relationship between CS and UCS that are necessary and sufficient for the CS to acquire the power to elicit the conditioned response, and a consensus appears to be emerging that the crucial factor is the extent to which the CS is a predictor of the UCS (Rescorla, 1967). However, the notion of predictiveness does not appear to differ from *relative proximity* (of the CS to the UCS, or of a stimulus to reinforcement) which, as we have seen (Principle of Reinforcement 3, above), is a factor of wide applicability in operant conditioning. Thus, for all practical purposes, classical conditioning may be defined operationally as a class of reinforcement schedules that involve presentation of reinforcement independently of the subject's behavior.

We conclude, therefore, that the division of the field of learning into two classes—classical and instrumental conditioning—each governed by separate sets of principles, has no basis in fact. As an alternative, we suggest an analysis based on the principles of behavioral variation and reinforcement we have already discussed. In terms of this analysis, all adaptive behavior is subsumed under an expanded version of the Law of Effect, and a given situation is to be understood in terms of two factors: (*a*) the reinforcement schedule, that is, the rule prescribing the delivery of reinforcement, or, more generally, stimuli, in relation to the behavior of the organism, and (*b*) the nature of the response under consideration. In terms of such an analysis, the properties normally considered as distinctive of classical conditioning may, once attention is directed to the question, be seen as due in part to a reinforcement schedule that happens to prescribe no correlation between the delivery of reinforcement and the subject's behavior, and in part to the special properties of responses such as salivation (see Implications 2 and 3, below).

Implications. Several puzzles become clearer once classical and instrumental conditioning are no longer regarded as separate processes:

1. In the earlier discussion of auto-shaped pecking, Williams and Williams noted that "the directed quality of the induced pecking does not follow naturally from respondent principles." Since we question this framework as a general description of anything outside the salivation experiment, its inability to deal with this particular situation poses no problem. The appearance of key pecking in the auto-shaping and superstition situations is, of course, not fully understood. It may reflect a special susceptibility of consummatory responses to the principle of stimulus substitution, as suggested earlier, or the action of transfer principles in ways that reflect something in common among the past histories of most pigeons. Similarly, little can be said about the directedness of pecking until the conditions under which this response is, and is not, directed have been more fully explored; although the results so far (with pigeons and monkeys) suggest that skeletal responses may always

be directed. Rather than attempt to explain (or explain away) these characteristics, the best course seems to be simply to observe and classify these and other behaviors under a variety of response-independent and response-dependent schedules, in the hope that hypotheses can be devised that have more hope of generality than anything that can be inferred from the scanty data presently available. We return to these issues in the concluding section.

2. There is little doubt that most autonomic responses are more easily conditioned by response-independent schedules than by response-dependent ones. However, this need not imply the existence of two different kinds of conditioning. It can as well be interpreted as reflecting the existence of internal controlling factors that are not amenable to the principles of reinforcement. Because of these fixed factors, these responses are not as free to come under the control of external stimulus factors as are skeletal responses that perform no function in the internal economy of the organism. Moreover, this principle can be extended beyond autonomic responses to deal with any response that is strongly connected to any stimulus, internal or external. Thus, in a recent discussion of operant conditioning of drinking, Black (1970) writes:

> This discussion suggests that one dimension along which responses might be classified with respect to operant conditioning is the degree to which they are constrained from being changed by operant reinforcement by the properties of the neural subsystems of which they are a part. The regulatory systems can vary from very simple reflexes, such as the knee-jerk, to complex instinctive ones, such as those involved in courtship. The main point is not the complexity of the subsystems but rather the extent to which they limit the conditions under which operant reinforcement will work [p. 267].

Segal (1970) has made a similar suggestion in a thoughtful discussion dealing with a number of the points raised in the present paper.

As an example, the difficulties associated with demonstrating operant conditioning of heart rate can be viewed as analogous to the problem of acquiring control over an operant response that is already under the control of other variables, that is, such schedules are really concurrent rather than simple schedules. The heart rate problem is perhaps analogous to training an animal to alter his rate of bar pressing to receive food, while pressing the same bar is also necessary to obtain another reinforcer (such as oxygen or heat). Even without considering the problem of interactions among drives, one would not be surprised to find rather weak control by the food reinforcer.

3. Williams (1965) recorded salivation in dogs while they were bar pressing for food reinforcement on both fixed-interval and fixed-ratio reinforcement schedules. He found that the onset of salivation within each interreinforcement interval approximately coincided with the onset of bar pressing in the fixed-interval case, but began *later* than bar pressing on fixed-ratio. This interesting result is incompatible with an explanation of the operant response in terms of an underlying classical conditioning process. However, it may be understood in terms of the view we have been proposing by assuming that (*a*) the occurrence of each response (bar pressing or salivation) is separately and independently determined by the conditions of reinforcement peculiar to it, and (*b*) both responses tend to occur at times of greatest relative proximity to reinforcement (Principle of Reinforcement 3). In the fixed-interval case, these two assumptions predict a similar time of occurrence for both behaviors, because time since the beginning of a trial is, for both responses, the best predictor of reinforcement. However, in the fixed-ratio case, it is apparent that no matter what determines the time of onset of bar pressing, once it has stabilized its onset provides a better predictor of reinforcement than does trial time (since the fixed number of responses making up the fixed ratio take an approximately fixed time). Thus, at asymptote our two assumptions imply that salivation should be reliably delayed with respect to bar pressing on fixed-ratio, but not on fixed-interval, as Williams reports.

However, there is no reason to expect this delay early in training, since the animal is not in a position to learn the cue significance of bar pressing until it has more or less stabilized. This expectation is also confirmed by Williams (1965), who notes that the delayed onset of salivation in the fixed-ratio case "emerged only after repeated exposure to the schedules [pp. 344–345]."

On the basis of a failure to find salivation *preceding* bar pressing on a spaced-responding ("controlled latency") procedure, Williams (1965) concludes that "the hypothesis that the two measures are independent may be rejected [p. 347]," which contradicts Assumption *a*, above. However, since *both* responses are assumed (Assumption *b*) to occur at times of greatest relative proximity to reinforcement, and reinforcement *cannot* occur *before* a bar press if it is always contingent upon bar pressing, there is no reason to expect salivation reliably to precede bar pressing under conditions where the common reinforcement for both responses depends on bar pressing alone. Thus, the apparent asymmetry between salivation and bar pressing observed by Williams may simply reflect an asymmetry between the conditions of reinforcement for each response, and need not imply any fixed internal linkage between them.

4. Recently, considerable attention has been devoted to avoidance of particular foods conditioned by a nauseous experience (induced by insulin or X rays) taking place several hours after ingestion (Garcia, Ervin, & Koelling, 1966; Kalat & Rozin, 1970; Revusky & Bedarf, 1967; Rozin, 1969). Since the CS–UCS interval in these experiments is considerably longer than is customary in classical conditioning experiments, these data are even less congenial to a Pavlovian analysis than the auto-shaping results. They hint at the existence of a number of unsuspected built-in linkages between response systems and various salient stimuli. Such linkages are not unexpected rom a broad evolutionary point of view that sees principles of variation and reinforcenent as behavioral characteristics that are eparately selected for, and bear as much or as little) relationship to one another as o morphological characteristics.

Difficulties of the Proposed Classification

Science is conservative and, quite correctly, resists most attempts to alter an established theoretical framework. We have already tried to show that the number of anomalies facing current learning theories is sufficient to justify a search for alternatives. Nevertheless, the radical appearance of the scheme we suggest is a substantial obstacle to its consideration. It is important, therefore, to point out that it is little more than an extension and reorganization of familiar concepts, that is, reinforcement, S–R behavior units, learning principles such as transfer, and ethological observations on species-related behaviors. The difference is therefore largely one of emphasis and selection rather than the introduction of wholly novel ideas.

Any discussion of evolution and learning naturally brings to mind the learning-instinct issue. There is no simple parallel between this dichotomy and anything in the scheme we propose. The origin of every behavior is supposed traceable to principles of variation; if, for example, in a particular case a principle of transfer is involved, one might want to say that the behavior is learned. However, the question must then simply be asked again about the previous situation from which transfer has supposedly occurred. In this way, almost any question about the relative roles of heredity and environment will involve unraveling the whole of ontogeny. This conclusion will not be unfamiliar to ethologists (cf. Beach, 1955).

It is also important to emphasize that we have not been directly concerned with the evolution of the capacity to learn; although it may be that increasing knowledge of variation will shed light on this issue.

One objection that may be raised to the proposed scheme is that it is derived from and deals explicitly only with positive reinforcement. However, a recent account of behavior sustained by negative reinforcement (Bolles, 1970) is in perfect agreement with our position. Bolles points out that some activities are much more easily conditioned than others in avoidance situations, and these are the unconditioned activities that normally occur in a variety of potentially dangerous circumstances. These species-specific defense reactions, in Bolles' terminology, occur in advance of reinforcement (i.e., the avoidance of electric shock) —in our terms, they are determined initially by one of the principles of behavioral variation. The lack of arbitrariness of the response is perhaps more obvious in avoidance than in any other situation because of the complexity of the schedules involved: the animal must usually learn something about the pattern of occurrence of an intermittent aversive event, in the absence of responding, before he is in a position to detect alterations in that pattern correlated with his own behavior. Although a similar situation prevails in all reinforcement schedules, the change to be discriminated seems considerably easier both in appetitive conditioning, where the shift is from zero reinforcement in the absence of responding to reinforcement following every response, and in escape, where it is from continuous presence of the aversive stimulus in the absence of responding to complete absence following each response. Bolles suggests other reasons, related to the limited opportunities for avoidance (in the schedule sense) in the wild life of small mammals, and thus the limited opportunities for the capacity to avoid to be selected in phylogeny.

The strongest point in favor of our proposal is its promise of parsimony. Consequently, the most damaging criticism that can be directed against it is the absence of firm specification of the principles of reinforcement and variation. This appears to allow the creation of such principles at will, enabling us to explain everything—and nothing. There are two defenses against this criticism. First, we again emphasize the tentative nature of the principles we have suggested. The overlap among the principles of variation, particularly, suggests that our list is provisional. Second, there is the strong possibility that clear recognition of the distinction between variation and reinforcement may be essential to further advance. In defense of this proposition, we first briefly discuss some examples from synoptic accounts of current learning theory, which show it to incorporate few safeguards against multiple explanations for phenomena. Since our scheme of classification is at least internally consistent and forces one to relate each new principle of variation to others that already exist, it has some advantages in this respect. Second, we discuss the controversy between cognitive and behavioristic theorists regarding the role of structure in behavior, in relation to a similar controversy in the history of evolutionary thought. The persistence of this controversy and its amenability to analysis in terms of variation and reinforcement suggest that our classification may be of some value despite its incompleteness.

1. Hilgard and Marquis (1940) list as principles of reinforcement: stimulus substitution, expectancy, and the principle (law) of effect. It should be apparent from the earlier arguments that the Law of Effect is the result of the combined effect of both variation and reinforcement, stimulus substitution is a principle of variation, and expectancy refers to a general characteristic which can be imputed to most learning. Consequently this set of terms allows for considerable uncertainty in application to particular situations. For example, our analysis of the Williams and Williams experiment (see p. 33) makes use of both stimulus substitution and a principle of reinforcement analogous to what Hilgard and Marquis mean by the Law of Effect. Yet the same situation could also be analyzed in terms of expectancy; and it appears to be incompatible with the Law of Effect as traditionally understood. Progress since 1940 has not been dramatic, as illustrated by a list of "elementary conditioning processes" inventoried by Jenkins (1970) in connection with his work on cyclic reinforcement schedules: generalization, delay of reinforcement, conditioned reinforcement, unconditioned effects of eating, frustration effects, effects related to "behavioral contrast." Despite the number of these processes and the lack of any obvious relationship among them, Jenkins finds that they are unable to account for some rather simple features of his data, which require a description in terms of the relative proximity

of a stimulus or a response to reinforcement as a major determiner (see Principles of Reinforcement, above).

2. There is a history of fruitless controversy between many behaviorists, who place little emphasis on the structural properties of behavior, and students of cognitive processes, who see structure as the most interesting and important behavioral attribute (cf. Staddon, 1967, 1969, for a discussion in relation to operant conditioning). The distinction between variation and reinforcement can shed some light on this issue, which can be illustrated by briefly considering the contrasting views of Skinner and Chomsky (Chomsky, 1959; MacCorquodale, 1969) on the causation of learned behavior.

Chomsky's major concern is with principles of variation, in our terms, as is clear from his emphasis on the rule-governed nature of language (see Principle of Behavioral Variation 4, above).

Skinner's position is less obvious, but becomes clear from his account of the shaping of behavior (Skinner, 1953); he writes:

> Operant conditioning shapes behavior as a sculptor shapes a lump of clay. Although at some point the sculptor seems to have produced an entirely novel object, we can always follow the process back to the original undifferentiated lump, and we can make the successive stages by which we return to this condition as small as we wish. At no point does anything emerge which is very different from what preceded it. The final product seems to have a special unity or integrity of design, but we cannot find a point at which this suddenly appears. In the same sense, an operant is not something which appears full grown in the behavior of the organism. It is the result of a continuous shaping process [p. 91].

For Skinner, apparently, moment-to-moment variation in behavior is small in magnitude, and essentially random (in the sense that it is unrelated to the final goal) in direction. Behavior is the result of the "accumulation . . . of indefinite variations which have proved serviceable" in Darwin's phrase. The similarity to natural selection is further emphasized by Darwin's (1951) account of the evolution of complex structures:

> If it could be demonstrated that any complex organ existed, which could not possibly have been formed by numerous, successive, slight modifications, my theory would absolutely break down [p. 191].

In the history of evolution after Darwin, the rediscovery of Mendel's laws led to a retreat from gradualism in favor of a saltationism that traced evolutionary progress (especially evolutionary novelty) to large changes (mutations) of a more purposive sort (cf. Mayr, 1960). This position is closer to the view of Chomsky and other cognitive theorists, who tend to stress the importance of insight and other rules of composition that can produce sudden jumps in behavior.

The history of evolution has not supported the saltationist view. Fisher (1930) showed that large changes are much less likely to be adaptive than small ones, and Haldane and others have shown by a variety of arguments that the time available for evolution by the selection of small variations is more than sufficient to account for the observed differences among taxa: "The saltationism of the early Mendelians has been refuted in all its aspects [Mayr, 1960, p. 350]."

At a superficial level, therefore, these comparisons might appear to favor Skinner's gradualism and emphasis on reinforcement (selection), to the detriment of the cognitive position. This is probably unjustifiable for two main reasons. First, detailed analysis of complex problem solving clearly indicates the insufficiency of random variation as an account of the process (e.g., Neisser, 1967). The heuristics that are employed may, of course, be attributed to past learning based entirely on random variation. However, this suggestion meets with quantitative difficulties when applied to the development of language—the best-studied example of rule-governed behavior. Although calculations in this area are of limited validity in the absence of established principles of variation (analogous to Mendelian genetics), the attempts that have been made seem to indicate that the time available in ontogeny for the development of language is incompatible with any kind of learning by random variation (Chomsky, 1962; McNeill, 1968). This negative result is the opposite of Haldane's affirmative conclusion on the sufficiency of small mutations as a basis for phylogenetic changes. It suggests that neither the heuristics employed in complex problem solving, nor the rules of syntax, need to be built up entirely de novo during ontogeny.

Second, the relationship between evolution and learning is such as to allow greater flexibility to learning. This is because natural selection can only be a response to small differences in "fitness": the most fit genotype will tend to prevail, although the species as a whole may thereby be led into an evolutionary blind alley. In terms of contemporary accounts of goal-directed mechanisms, natural selection represents a hill-climbing process (Minsky, 1961) and has no provision for prediction. This is clear in a familiar analogy due to Sewall Wright (1931) which shows the relationships among selection, structure, and variation. He pictures the field of possible structural variation as a landscape with hills and valleys. The range of variation present in a population of organisms is represented by a closed area on this landscape. Selection pressure is represented by the gradient (upward slope) of the landscape, so that each *peak* is an adaptive optimum for a given constellation of characters; valleys represent unstable equilibria yielding so-called centrifugal selection. If the area representing a given species includes a single adaptive peak, selection will be centripetal, so that the species will tend to cluster more and more closely around the peak. Thus, small mutations are more likely to lead to improvements in fitness than large ones (with a limiting probability of .5, as Fisher, 1930, has shown), and the consequent predominant role of such small differences in fitness in the evolutionary process becomes obvious.

Learning is not so limited, however, because the principles of variation can be weighted to take account of regularities in the past history of the species (these are Skinner's, 1966b, "phylogenic contingencies"); that is, behavior need not occur at random in advance of reinforcement, but can reflect a priori probabilities that have been selected for during phylogeny. Other more complex strategies of this sort may also be built up by natural selection, giving learning a predictive capacity largely denied to evolution itself. Thus, although learned behavior reflects differences in reinforcement rate, just as evolution reflects selection on the basis of relative fitness, it need *not* be generally true either that small changes in behavior are more likely to be adaptive than large ones, or that the direction of change is unrelated to the final goal—as Skinner's account implies, and as is usually (although not invariably) the case in evolution. However, since the more elaborate principles of variation must themselves be built up step-by-step by natural selection, it is to be expected that the pattern and range of behavioral variation must bear some relationship to the phylogenetic status of the organism: "higher" organisms, such as man, are likely to have developed more complex principles of variation than "lower" organisms, such as the pigeon.

Thus, the major focus of the argument between Skinner and Chomsky is not on the importance of reinforcement, but about the complexity of the principles of variation that determine the nature of behavior in advance of reinforcement. Since Skinner derives his ideas from work on rats and pigeons, and Chomsky from the study of human language, there are considerable grounds for a disagreement, especially if its basis is not clearly perceived by either party. Clear conceptual separation of variation from reinforcement makes this kind of confusion much less likely.

CONCLUSION

The argument so far has served to draw attention to a number of generalizations about steady-state conditioning situations:

1. Most such situations involve some times and stimuli associated with relatively high reinforcement probability (e.g., the period at the end of the interval on fixed-interval schedules), and others associated with relatively low reinforcement probability (e.g., the period at the beginning of the interval).

2. The *terminal response* (a discriminated operant in Skinner's terminology) is restricted to periods of relatively high reinforcement probability. This distribution of the terminal response with respect to time

and stimuli corresponds to a principle of reinforcement that relates the strength of a response to the relative frequency, density, or proximity of reinforcement associated with that response (Catania & Reynolds, 1968; Herrnstein, 1970; Jenkins, 1970).

3. The type, as opposed to the temporal and stimulus location, of the terminal response in situations involving both response-dependent and response-independent reinforcement is determined by the interaction between principles of variation (e.g., transfer, stimulus substitution) that describe the occurrence of the response, in advance of reinforcement, and principles of reinforcement that determine whether it will persist or not (selective function of reinforcement).

4. Periods of low reinforcement probability are generally associated with *interim activities,* resembling appetitive behavior. If appropriate stimuli (goal objects) are provided, stereotyped *adjunctive behavior* (e.g., polydipsia, pica; Falk, 1969) takes the place of the more variable and relatively undirected interim activities.

5. Both terminal response and interim activities are more correctly labeled as predisposing conditions or *states* rather than behaviors, since in the absence of response dependency, the type of activity falling into these categories is not fixed. Thus, drinking, wheel running, fighting, pecking, and a number of other activities may be either terminal or adjunctive behaviors, depending on historical and stimulus factors (Segal, 1969b; Skinner, 1959; Skinner & Morse, 1958). Directing factors for adjunctive behavior are the availability of appropriate goal objects (see above), and factors related to reinforcement that render some kinds of activity more probable than others: for example, polydipsia appears to partially displace both adjunctive wheel-running (Segal 1969a) and chewing-manipulatory behavior (Freed & Hymowitz, 1969) in rats (in situations with food as terminal reinforcer), even when both supporting stimuli are concurrently available.

The Terminal Response

We have already discussed the probable role of principles of variation such as stimulus substitution in describing the origin of terminal responses such as Pecking and Head in magazine in the superstition situation. Before turning to the more complex matter of the interim activities, a word should be said about the paradoxical results of the Williams and Williams (1969) study, in which they found persistent pecking at a brief stimulus, ending in reinforcement, despite the fact that a key peck terminated the stimulus and thus prevented reinforcement. In terms of our analysis, this situation pits reinforcement and variation against one another; thus, (*a*) the predictable delivery of food at the end of each presentation of a brief key stimulus may be a sufficient condition for the occurrence of key pecking in that stimulus, by the principle of stimulus substitution. But, (*b*) because of the response-contingency, the occurrence of a peck turns off the stimulus, omitting reinforcement on that occasion, and thus reducing the overall reinforcement rate. In turn, this reduction in reinforcement rate will, via Principle of Reinforcement 3, tend to reduce the tendency to make any terminal response, including pecking, in that situation. This process will continue until the tendency to peck has been sufficiently reduced to allow the key stimulus to continue unpecked until the delivery of reinforcement, which will again provide the occasion for the operation of the stimulus substitution principle, making pecking likely once again. Thus, an equilibrium will be established at a rate of pecking higher than zero, but less than the rate which would obtain if pecking had no effect on reinforcement rate. Extinction of pecking takes place if key pecking prevents reinforcement, but does not turn off the key stimulus, because the predictability of reinforcement, and thus the necessary condition for the operation of stimulus substitution, is thereby destroyed. Under these conditions, variation and reinforcement combine to weaken the tendency to peck, which therefore declines relatively rapidly.

The data on "instinctive drift" reported by Breland and Breland (1961, 1966) are also compatible with this kind of analysis.[7] However, their results strongly suggest that too much stress should not be laid on the apparent identity we (and others, e.g., Brown & Jenkins, 1968; Williams & Williams, 1969; Wolin, 1968) have found between the topography of the terminal response and the (unconditioned) response made to the terminal reinforcer. As in the case of adjunctive behavior (discussed below) and behavior elicited by central stimulation in reinforcing brain areas (e.g., Glickman & Schiff, 1967; von Holst & von Saint Paul, 1963), the effect of the experimental procedures we have described appears to be the induction of a *state* (in our terminology) or "mood" (von Holst & von Saint Paul) which makes some kinds of activity much more likely than others but preserves some flexibility in the animal's mode of response to the stimulating environment.

For example, Breland and Breland (1961) describe two situations, both showing "instinctive drift" in chickens, but only one of which conforms to the stimulus substitution principle. In the first case, the chicken pecked a ball which he had learned to project (via a remote firing mechanism) at a target—a hit being immediately followed by food. The history of contiguity between the moving ball and reinforcement and the similarity of the responses to both ball and food (i.e., pecking) fit easily into the stimulus substitution paradigm. In the second case, however, the chicken was reinforced for a chain of responses, the last of which involved standing on a platform for 15 seconds. After training, the chicken showed vigorous ground scratching while standing on the platform. This response has no topographic resemblance to pecking food, but is, of course, a universal food-getting behavior in chickens and typically occurs in the vicinity of food.

These examples, and others discussed by Breland and Breland (e.g., raccoons "washing" poker chips, porpoises swallowing manipulanda, pigs "rooting" tokens, etc.), as well as exceptions to stimulus substitution in the auto-shaping literature (e.g., Rachlin, 1969; Sidman & Fletcher, 1968), are compatible with a more general notion, to the effect that the stimulus (temporal or exteroceptive) most predictive of reinforcement comes to control a state or mood (the terminal state) appropriate to that reinforcer. The particular activity which occurs during the terminal period will then depend on principles of variation, which take into account *both* its motivational properties (e.g., food-related activities become more likely if the terminal reinforcer is food), *and* the nature of the stimulating environment; that is, the nature of the terminal state determines what stimuli will be effective in eliciting what behavior. Thus, the Breland and Breland chickens pecked when the stimulus defining the terminal state was appropriate (in some sense) for pecking, but scratched when it was not; pigeons peck the key in auto-shaping experiments, but (based on our results) are slower to peck in the superstition situation—presumably because an appropriate target is not provided. Similarly, Glickman and Schiff (1967) reviewed a large number of studies of behavior induced by direct brain stimulation which suggested that the effect of stimulation at a particular site is to induce a predisposing motivational condition or state which may lead to a variety of behaviors depending on the presence or absence of appropriate supporting stimuli. More recent work (e.g., Valenstein, Cox, & Kakolewski, 1970) further emphasizes the similarity between the terminal state and behavior induced by central stimulation:

> Hypothalamic stimulation does not activate only one specific behavior pattern. The stimulation seems to excite the substrate for a group of responses that in a given species are related to a common state [Valenstein et al., 1970, p. 30].

---

[7] The delayed appearance of the food-related behaviors in the Breland and Breland situations (as compared to the relatively rapid emergence of pecking in auto-shaping experiments) reflects the fact that food delivery could not become predictable (the necessary condition for the operation of the stimulus substitution principle) until after the animals had learned to produce it by making the required "arbitrary" response. However, once this initial response was learned (via principles of variation other than stimulus substitution), the stage was set for the operation of stimulus substitution, which could then override the original learning.

Within this more general framework, stimulus substitution becomes a special case, which simply reflects the fact that the response normally made to the terminal reinforcer often becomes highly probable when the animal is in a terminal state corresponding to that reinforcer, and will occur if minimal environmental support is provided.

Interim Activities and Adjunctive Behavior

The principles of variation and reinforcement so far discussed refer to the origin and maintenance of the terminal response. The interim activities (including adjunctive behavior) require a separate although complementary account, to which we now turn.

There is as yet no general agreement on the causal factors underlying adjunctive behavior (i.e., behavior occurring during the interim period on a variety of intermittent reinforcement schedules). As we have already noted (p. 13), data presently available appear to rule out simple physiological interpretations (Falk, 1969), although they do not clearly point to any alternative. Some help is offered by the similarities between adjunctive behavior and displacement activities: the same explanation should be adequate for both. Possibilities are also somewhat restricted by general functional considerations. In the present section, we state a tentative general hypothesis and its empirical basis in what is presently known of adjunctive behavior. The relationship of this hypothesis to accounts of displacement behavior and to general adaptiveness is discussed in the following section.

The hypothesis may be stated in the form of three propositions:

1. The interim and terminal periods correspond to states, in the sense described earlier, defined by (*a*) the class of reinforcer or reinforcers that are effective at that time, and (*b*) the applicability of principles of variation appropriate to that class of reinforcer (e.g., food-related behaviors are likely to occur during the food state, defense reactions during the fear state, etc.).

2. The terminal state corresponds to the terminal reinforcer; the state during the interim period corresponds to *all other reinforcers*, although all need not be equally effective. The linkage between terminal and interim states is assumed to be direct and reciprocal, so that the strength (defined below in terms of *rate*) of activities during the interim period is *directly related* to the strength of the terminal response.

3. The strength of the terminal response is directly related to the "value" of the reinforcement schedule; that is, to relative rate and amount of reinforcement and to motivational factors (e.g., deprivation).

As a consequence of the reciprocal interaction between the terminal and interim states (Proposition 2) and the dependence of the terminal response on the value of the reinforcement schedule (Proposition 3), the strength of behaviors associated with the interim period will be determined both by the value of the terminal reinforcement schedule, as well as by the value of the reinforcers proper to them.

The notion of the interim and terminal periods as states (Proposition 1) has already been discussed. The dependence of the strength of the terminal response on variables related to the value of the reinforcement schedule (Proposition 3) should also encounter no opposition. It remains to show, first, that reinforcers other than the terminal reinforcer are effective during the interim period; second, that a number of different reinforcers may be effective at this time; and third, that the direct relation between the strength of the terminal response and the strength of adjunctive behavior implied by Propositions 2 and 3 has some basis in fact. Evidence for the effectiveness of reinforcers other than the terminal reinforcer, during the interim period, comes largely from studies of polydipsia (excessive drinking), induced by intermittent schedules of food reinforcement, as follows: (*a*) Consummatory behavior (e.g., drinking) occurs in the presence of the appropriate goal object (water) during the interim period on a variety of schedules. (*b*) This goal object can reinforce operant behavior:

> If water is not freely available . . . concurrently with a food schedule, but is available in small portions contingent upon the completion of a fixed-ratio schedule, polydipsia is acquired and will sustain large fixed-ratios [Falk, 1970, p. 297].

Azrin (1964, reported in Falk, 1970) has found similar schedule control, in pigeons, by a bird provided as a target for schedule-induced aggression. (*c*) Polydipsic drinking can be both increased and decreased by appropriate alteration of the palatability of the liquid available (Falk, 1964, 1966). (*d*) Polydipsia is usually reduced by using liquid terminal reinforcers. In the cases where the liquid used is at least as reinforcing to a hungry rat as dry food pellets (e.g., liquid Metrecal, condensed milk, liquid monkey diet; Falk, 1964, 1966; Hawkins, Everett, Githens, & Schrot, 1970; Stein, 1964), this decrease may be attributed to a direct effect on the thirst system (state), due to the water content, as in water preloading (see below). Wesson oil as terminal reinforcer is also less effective than food pellets in producing polydipsia (Stricker & Adair, 1966), although it contains no water, but is probably also less reinforcing and may reduce polydipsia for this reason (see Propositions 2 and 3). (*e*) Acquisition of polydipsia can be prevented by presession stomach loads of water (Chapman, 1969, reported in Falk, 1970), although established polydipsia is little affected. This kind of effect is also found with food-motivated terminal responses which, once established, will continue to occur at a somewhat reduced rate even in the presence of ad lib food (Neuringer, 1970a).

That more than one reinforcer is effective during the interim period is suggested by the facts that (*a*) the interim activities (which occur in the absence of appropriate goal objects) have no consistent direction and are not obviously related to any particular reinforcer; (*b*) a variety of goal objects—water, wood shavings, another animal—are sufficient to elicit appropriate consummatory reactions (drinking, chewing or eating, aggression); (*c*) physiological data, suggesting short-term reciprocal interactions between hunger and thirst drives which were induced centrally by electrical or chemical stimulation (Grossman, 1962; von Holst & von Saint Paul, 1963), indicate a possible mechanism for the simultaneous effectiveness of reinforcers other than food at nonfood times (interim periods) and food at food times (terminal periods). While it is as well to be cautious in generalizing both across species and from physiology to behavior, these hypothalamic mechanisms are evidently quite similar in birds and mammals (cf. Åkerman, Andersson, Fabricius, & Svensson, 1960), and recent work supports the similarity between schedule-induced drinking and drinking induced by direct hypothalamic stimulation implied by these comparisons (Burks & Fisher, 1970). This kind of reciprocal interaction suggests that at a time when activity motivated by hunger is suppressed (e.g., during the period of interim activities on a food schedule), activities motivated by thirst might be facilitated.

That the effective reinforcers during the interim period are other than the terminal reinforcer is suggested both by the nonoccurrence of the terminal response at that time and by data reported by Segal (1969b) showing a shift in the status of drinking—from an adjunctive behavior, occurring early in the interval, to a terminal response, occurring largely at the end. The changeover took several experimental sessions, but it suggests that in the steady state, the same activity is unlikely to occur during both terminal and interim periods, as might be expected if these periods are associated with the action of different reinforcers.

Propositions 2 and 3 in combination imply that the strength of adjunctive behaviors, like that of the terminal response, should be directly related to the "value" of the reinforcement schedule, as indexed by motivational and reinforcement variables. The evidence in favor of this deduction is as follows: (*a*) With the interreinforcement interval held constant, the amount of polydipsic drinking in a session of fixed length is inversely related to body weight (Falk, 1969). Similar results have been reported for schedule-induced attack and air licking (Falk, 1970). (*b*) Postpellet pause (until the onset of polydipsic drinking) increases as a function of the interval duration (Segal et al., 1965). (*c*) Rate of licking within a drink bout tends to decrease as a function of interval length (Segal et al., 1965). (*d*) Rate of polydipsic drinking is increased by increasing the size of food reinforcement

(Hawkins et al., 1970). Hawkins et al. attribute a contrary result by Falk (1967) to session length differences between the one- and two-pellet conditions of Falk's experiment. (*e*) Polydipsic drinking falls off drastically at fixed-interval values longer than about 3 minutes (Falk, 1966; Segal et al., 1965). Falk (1966) also reports a *direct* relationship between polydipsia (measured as total amount drunk per session) and fixed-interval length over the range 2–180 seconds. However, based on the results of experiments indicating the relative constancy of ingestion rate within a drinking bout in rats (Davis & Keehn, 1959; Schaeffer & Premack, 1961; Stellar & Hill, 1952), and the different times available for polydipsic drinking under different fixed-interval values, Falk's finding of a direct relationship between fixed-interval value and total amount drunk, over part of the range, is compatible with an overall *inverse* relationship in terms of *rate* of drinking. This inference is confirmed by recent data reported by Hawkins et al. (1970) which show a monotonically decreasing ingestion rate as a function of fixed-interval length, over the range 1–5 minutes. Given that overall ingestion rate is probably better than total amount drunk as a measure of the tendency to drink, this finding is both consistent with the deduction from Propositions 2 and 3 and more easily reconciled with other measures that indicate an inverse relationship between tendency to drink and frequency of reinforcement. Falk's finding that total amount drunk is maximal at intermediate interval values may reflect, therefore, an optimal balance between two factors: tendency to drink, which decreases as interval value increases, and time available for drinking, which increases with interval value.

The above account, in terms of interaction between motivational systems (states), seems to be the simplest that can presently be given of both adjunctive behavior and displacement activities. Its relationships to these activities, to the ethological interpretation of them, and to general functional considerations are discussed in the next section.

Adjunctive Behavior and Displacement Activities

Most behavioral and morphological characteristics are adaptive, in the sense that they can be directly related to the reproductive fitness of the organism in its natural environment. Morphological exceptions to this rule are either the result of "correlated variation," in Darwin's phrase, or are vestigial characters, in the process of being lost. In neither case do they show the ubiquity and reliability that distinguish adjunctive and displacement activities. It is very likely, therefore, that these behaviors reflect a function, or functions, of considerable adaptive value to animals in the wild. What might this function be?

Learning theories generally consider the organism to be motivated by one thing at a time, for example, hunger, thirst, exploratory drive, etc. In like spirit, most learning experiments are designed to ensure the predominance of one kind of motivation at the expense of all others. In the wild, however, animals must allocate their time among a variety of activities so as to both satisfy current wants and anticipate future ones. It is reasonable to assume that there has been considerable selection pressure favoring an optimal balance among the various possibilities.

We have already noted the fact, related to stimulus discrimination, that animals tend to make the terminal response only at times when reinforcement is likely. This fact, and the principle of reinforcement based on it, might seem to reflect some kind of Law of Effort, since responding at times or places when reinforcement never occurs is obviously wasteful. However, a considerable weight of evidence suggests that the Law of Effort is not a major psychological principle, since it is easy to devise situations in which animals make many more responses than necessary (cf. Ferster & Skinner, 1957). A Law of Effort principle would also not explain active *avoidance* of situations associated with nonreinforcement.

A more plausible alternative is that these facts are related to animals' need to budget their time effectively. In these terms, a time, or stimulus, reliably associated with the absence of a given reinforcer provides information just as useful as a time perfectly correlated with the delivery of that reinforcer, since it permits the animal to attend to present and future needs other than the one associated with the absent reinforcer. However, other potentialities of the environment cannot usually be sampled as long as the animal remains in the vicinity of the unavailable reinforcer. One might expect, therefore, that natural selection will have fostered the development of a mechanism to ensure that animals avoid places at times when, on the basis of past experience, they have learned that reinforcement is not forthcoming.

Evolution is notoriously opportunistic in the sense that adaptation is achieved by whatever structural or functional means happen to be available. In the present case, we suggest that the means for ensuring that animals will not linger in the vicinity of food (or other reinforcers) at times when it is not available may be provided by the facilitation of drives other than the blocked one (Propositions 2 and 3); that is, that the relative aversiveness of the stimuli in the vicinity of food, during the interim period, may be a *direct effect* of the simultaneous suppression of the food state and facilitation of states associated with other reinforcers. In the wild, such facilitation will usually ensure that the animal leaves the situation to seek other reinforcers. Moreover, once the animal has left the situation, generalization decrement will ensure that the effect of factors acting to facilitate these other drives is reduced, restoring the animal to a state appropriate to his condition of deprivation and allowing him to take advantage of new opportunities to satisfy the previously blocked drive.

In experimental situations showing adjunctive behavior, however, the animal is *kept in the vicinity of the withheld reinforcer,* both by the physical restraint of the enclosure and, perhaps more importantly, by the properties of the reinforcement schedule. Since enclosure size has not been explicitly investigated, one cannot be sure of the relative importance of these two factors. In our laboratory, we have often observed animals on time-based schedules and noticed that pigeons tend to turn away from the key during the no-pecking phase of schedules such as the fixed-interval, which yield a period of no pecking followed by pecking. Figure 1, which shows the birds' orientation (R_1) as a function of postfood time, is a quantitative record of this effect. However, the avoidance of the key is much more complete on schedules which require pecking followed by no pecking (temporal go–no go schedules, Staddon, 1970a, 1970b), presumably because the no-pecking period is terminated by an external event (delivery of reinforcement) rather than by the bird returning to the key to peck. These observations provide some evidence both for the tendency of pigeons to avoid the key at times when key pecking is not reinforced and for the restriction placed on this tendency by fixed-interval schedules.

In situations involving external (rather than temporal) stimuli, there is also considerable evidence for the aversive character of stimuli associated with nonreinforcement when they occur in a context associated with reinforcement (e.g., multiple and concurrent reinforcement schedules; cf. Beale & Winton, 1970; Catania, 1969; Terrace, 1966).

Thus, the temporal locus of adjunctive (and interim) behavior coincides with a period when, by other measures, the situation is aversive to the animal so that he will withdraw from it if he can. Studies of schedule-induced escape have shown that animals will learn to make a response during the interim period on fixed-ratio schedules that has the effect of removing them from the situation, even though the frequency of the terminal reinforcement may thereby be reduced (e.g., Azrin, 1961; Thompson, 1964). The present argument suggests that these data may reflect a general property of interim periods, although experimental results with other schedules are presently lacking.

Falk (1969, 1970) has ably summarized the similarities between adjunctive behavior and displacement activities, which include the apparent "irrelevance" of both kinds of activity, their association with situations in which a strong drive is blocked, and their modifiability by available stimuli and conditions of deprivation other than the major drive (cf. Morris, 1954; Rowell, 1961; Sevenster, 1961). A study by McFarland (1965) concerning displacement pecking induced by preventing drinking in thirsty doves further emphasizes the "state" property of displacement behavior:

> The evidence for the view that this pecking belongs to the feeding system . . . [is] as follows: 1. Total time spent pecking is increased by the presence of grain. 2. Time spent pecking is partly replaced by time spent at a specific food getting activity, when the birds have previously been trained to obtain food in this way. 3. Food deprivation . . . increases the time spent pecking when grain is present, and this effect is counteracted by [prefeeding] [p. 298].

The interpretation of adjunctive behavior that we have presented is very similar to the disinhibition hypothesis concerning displacement activities first suggested by Andrew (1956). Hinde (1966) summarizes this view as follows:

> When mutual incompatibility prevents the appearance of those types of behaviour which would otherwise have the highest priority, patterns which would otherwise have been suppressed are permitted to appear [p. 279].

Displacement behavior is usually exhibited in approach-avoidance conflict situations (e.g., territory defense, birds returning to the nest after an alarm, etc.) when the animal is consequently *prevented from leaving the situation.* We have already seen that during the interim period on intermittent reinforcement schedules, animals are also restrained in the situation both by the enclosure and, probably, by the properties of the schedule. Both situations therefore meet the conditions necessary and sufficient, by our hypothesis, for the *elevation* (rather than merely disinhibition) of motivational states other than the blocked one, leading in the schedule case to adjunctive behavior, and in the approach-avoidance case to displacement, redirection, or vacuum activities—the particular activities being a function of the strength and nature of the blocked response, the proximity to the goal, the stimuli available, the past history of the animal, and the duration of the blocking. This modification causes no difficulty in application to displacement behavior and meets theoretical objections to certain forms of the disinhibition hypothesis raised by McFarland (1966).

Finally, our account of the adaptive significance of adjunctive behavior, as a reflection of the integrative capacities of the organism which enable it to strike an efficient balance among a number of activities, finds a counterpart in McFarland's (1966) account of the significance of displacement behavior:

> Thus it is suggested that the functional significance of displacement activities is that they are the by-product of a mechanism which enables animals to break away from a specific course of action, when progress in that course of action comes to a standstill [p. 231].

We conclude, therefore, that interim, adjunctive, and displacement behaviors may be grouped together on the basis of similar functional properties, similar probable causal factors, and similar adaptive role.

In summary, the argument relating to the interim period is as follows: (*a*) Extant data on adjunctive behavior are consistent with a tentative general interpretation in terms of interactions among motivational systems (states). (*b*) On the basis of general adaptive considerations, we have suggested the probable existence of a mechanism which enables animals to budget their time efficiently (e.g., by giving up temporarily ineffective activities). (*c*) McFarland has suggested that displacement and other "irrelevant" activities may reflect the action of such a mechanism. (*d*) Falk has pointed out the extensive similarities between adjunctive and displacement activities. (*e*) McFarland's suggestion may therefore be extended to interim and adjunctive behaviors, both on the basis of their resemblance to displacement activities and their restriction to the aversive interim period. Thus, the general interpretation of adjunctive behavior offered earlier gains additional support from the resemblances between adjunctive and displacement activities, from its similarity to the disinhibition hypothesis for displacement activities, and from its adequacy as a mechanism for enabling the animals to budget their time efficiently.

Epilogue

This ends our outline of conditioning. We have dealt with both terminal periods—which, we suggest, reflect a Law of Effect process that can best be understood by analogy with evolution by means of natural selection—and interim periods—which may reflect a mechanism enabling animals to allocate their activities efficiently. Learned behavior, under the relatively simple conditions of reinforcement schedules at least, is viewed as reflecting the sequencing, with respect to time and stimuli, of terminal and interim periods; and the scheme is therefore potentially comprehensive, although necessarily incomplete as to details.

Our proposal is founded on the belief that the most distinctive thing about living creatures is the balance they maintain among a number of tendencies to action, each one adaptive, yet each destructive if pursued to the exclusion of others. This emphasis on the *integration* of behavior has required that the scheme attempt to be comprehensive and that it relate in a natural way to biological and physiological considerations. Such merits as it possesses lie not in formal elegance or precision, but in an ability to organize otherwise unrelated facts and to suggest gaps where others may possibly be found.

REFERENCES

Åkerman, B., Andersson, E., Fabricius, E., & Svensson, L. Observations on central regulation of body temperature and of food and water intake in the pigeon (*Columba livia*). *Acta Physiologica, Scandinavica,* 1960, **50**, 328–336.

Andrew, R. J. Some remarks on behaviour in conflict situations, with special reference to *Emberiza* Spp. *British Journal of Animal Behaviour,* 1956, **4**, 41–45.

Antonitis, J. J. Response variability in the white rat during conditioning, extinction, and reconditioning. *Journal of Experimental Psychology,* 1951, **42**, 273–281.

Azrin, N. H. Time-out from positive reinforcement. *Science,* 1961, **133**, 382–383.

Azrin, N. H. Aggression. Paper presented at the meeting of the American Psychological Association, Los Angeles, September 1964.

Azrin, N. H., Hutchinson, R. R., & Hake, D. F. Extinction-induced aggression. *Journal of the Experimental Analysis of Behavior,* 1966, **9**, 191–204.

Beach, F. A. The descent of instinct. *Psychological Review,* 1955, **62**, 401–410.

Beale, I. L., & Winton, A. S. W. Inhibitory stimulus control in concurrent schedules. *Journal of the Experimental Analysis of Behavior,* 1970, **14**, 133–137.

Black, A. H. Constraints on the operant conditioning of drinking. In, *Schedule-induced and schedule-dependent phenomena.* Vol. 2. Toronto: Addiction Research Foundation, 1970.

Bolles, R. C. Species-specific defense reactions and avoidance learning. *Psychological Review,* 1970, **77**, 32–48.

Breland, K., & Breland, M. The misbehavior of organisms. *American Psychologist,* 1961, **16**, 661–664.

Breland, K., & Breland, M. *Animal behavior.* New York: Macmillan, 1966.

Broadbent, D. E. *Behaviour.* London: Methuen, 1961.

Brown, P. L., & Jenkins, H. M. Auto-shaping of the pigeon's key-peck. *Journal of the Experimental Analysis of Behavior,* 1968, **11**, 1–8.

Burks, C. D. Schedule-induced polydipsia: Are response-dependent schedules a limiting condition? *Journal of the Experimental Analysis of Behavior,* 1970, **13**, 351–358.

Burks, C. D., & Fisher, A. E. Anticholinergic blockade of schedule-induced polydipsia. *Physiology and Behavior,* 1970, **5**, 635–640.

Cane, V. Some ways of describing behaviour. In W. H. Thorpe & O. L. Zangwill (Eds.), *Current problems in animal behaviour.* London: Cambridge University Press, 1961.

Catania, A. C. Concurrent performances: Inhibition of one response by reinforcement of another. *Journal of the Experimental Analysis of Behavior,* 1969, **12**, 731–744.

Catania, A. C., & Reynolds, G. S. A quantitative analysis of the responding maintained by interval schedules of reinforcement. *Journal of the Experimental Analysis of Behavior,* 1968, **11**(Pt. 2), 327–383.

Chapman, H. W. Oropharyngeal determinants of non-regulatory drinking in the rat. Unpublished doctoral dissertation, University of Pennsylvania, 1969.

Chomsky, N. A review of B. F. Skinner's *Verbal behavior. Language,* 1959, **35**, 26–58.

Chomsky, N. Explanatory models in linguistics. In E. Nagel, P. Suppes, & A. Tarski (Eds.), *Logic, methodology and philosophy of science: Proceedings of the 1960 International Congress.* Stanford: Stanford University Press, 1962.

Craig, W. Appetites and aversions as constituents of instincts. *Biological Bulletin of the Marine Biological Laboratory,* Woods Hole, Mass., 1918, **XXXIV**, 91–107.

Darwin, C. *The origin of species.* Oxford: The University Press, 1951. (Reprinted from the sixth edition, 1872.)

Darwin, C. *The variation of animals and plants under domestication.* New York: Appleton, 1896. 2 vols.

DAVIS, J. D., & KEEHN, J. D. Magnitude of reinforcement and consummatory behavior. *Science,* 1959, **130**, 269–270.

ESTES, W. K. Of models and men. *American Psychologist,* 1957, **12**, 609–617.

FALK, J. L. Studies on schedule-induced polydipsia. In M. J. Wayner (Ed.), *Thirst: First international symposium on thirst in the regulation of body water.* New York: Pergamon Press, 1964.

FALK, J. L. Schedule-induced polydipsia as a function of fixed interval length. *Journal of the Experimental Analysis of Behavior,* 1966, **9**, 37–39.

FALK, J. L. Control of schedule-induced polydipsia: Type, size, and spacing of meals. *Journal of the Experimental Analysis of Behavior,* 1967, **10**, 199–206.

FALK, J. L. Conditions producing psychogenic polydipsia in animals. *Annals of the New York Academy of Sciences,* 1969, **157**, 569–593.

FALK, J. L. The nature and determinants of adjunctive behavior. In, *Schedule-induced and schedule-dependent phenomena.* Vol. 2. Toronto: Addiction Research Foundation, 1970.

FANTINO, E., & COLE, M. Sand-digging in mice: Functional autonomy? *Psychonomic Science,* 1968, **10**, 29–30.

FERSTER, C. B., & SKINNER, B. F. *Schedules of reinforcement.* New York: Appleton-Century-Crofts, 1957.

FISHER, R. A. *The genetical theory of natural selection.* Oxford: Clarendon Press, 1930.

FLORY, R. Attack behavior as a function of minimum inter-food interval. *Journal of the Experimental Analysis of Behavior,* 1969, **12**, 825–828.

FREED, E. X., & HYMOWITZ, N. A fortuitous observation regarding "psychogenic" polydipsia. *Psychological Reports,* 1969, **24**, 224–226.

GARCIA, J., ERVIN, F. R., & KOELLING, R. Learning with prolonged delay of reinforcement. *Psychonomic Science,* 1966, **5**, 121–122.

GHISELIN, M. T. *The triumph of the Darwinian method.* Berkeley: University of California Press, 1969.

GILBERT, R. M. Psychology and biology. *Canadian Psychologist,* 1970, **11**, 221–238.

GLICKMAN, S. E., & SCHIFF, B. B. A biological theory of reinforcement. *Psychological Review,* 1967, **74**, 81–109.

GLICKMAN, S. E., & SROGES, R. W. Curiosity in zoo animals. *Behaviour,* 1966, **26**, 151–188.

GONZALEZ, R. C., BEHREND, E. R., & BITTERMAN, M. E. Reversal learning and forgetting in bird and fish. *Science,* 1967, **158**, 519–521.

GROSSMAN, S. P. Direct adrenergic and cholinergic stimulation of hypothalamic mechanisms. *American Journal of Physiology,* 1962, **202**, 872–882.

GUTTMAN, N., & KALISH, H. I. Discriminability and stimulus generalization. *Journal of Experimental Psychology,* 1956, **51**, 79–88.

HALDANE, J. B. S. Natural selection. In P. R. Bell (Ed.), *Darwin's biological work.* Cambridge: University Press, 1959.

HARLOW, H. F. Learning set and error factor theory. In S. Koch (Ed.), *Psychology: A study of a science.* Vol. 2. New York: McGraw-Hill, 1959.

HAWKINS, T. D., EVERETT, P. B., GITHENS, S. H., & SCHROT, J. F. Adjunctive drinking: A functional analysis of water and alcohol ingestion. In, *Schedule-induced and schedule-dependent phenomena.* Vol. 1. Toronto: Addiction Research Foundation, 1970.

HERRNSTEIN, R. J. "Will." *Proceedings of the American Philosophical Society,* 1964, **108**, 455–458.

HERRNSTEIN, R. J. Superstition: A corollary of the principles of operant conditioning. In W. K. Honig (Ed.), *Operant behavior: Areas of research and application.* New York: Appleton-Century-Crofts, 1966.

HERRNSTEIN, R. J. On the law of effect. *Journal of the Experimental Analysis of Behavior,* 1970, **13**, 243–266.

HERRNSTEIN, R. J., & HINELINE, P. N. Negative reinforcement as shock-frequency reduction. *Journal of Experimental Analysis of Behavior,* 1966, **9**, 421–430.

HILGARD, E. R., & MARQUIS, D. G. *Conditioning and learning.* New York: Appleton-Century, 1940.

HINDE, R. A. *Animal behaviour: A synthesis of ethology and comparative psychology.* New York: McGraw-Hill, 1966.

HUTCHINSON, R. The production and maintenance of behavior by shock and shock-associated stimuli. In R. Ulrich (Chm.), The maintenance of responding through the presentation of electric shocks. Symposium presented at the meeting of the American Psychological Association, Miami, September 1970.

JENKINS, H. M. Sequential organization in schedules of reinforcement. In W. N. Schoenfeld & J. Farmer (Eds.), *Theory of reinforcement schedules.* New York: Appleton-Century-Crofts, 1970.

KALAT, J. W., & ROZIN, P. "Salience": A factor which can override temporal contiguity in taste-aversion learning. *Journal of Comparative and Physiological Psychology,* 1970, **71**, 192–197.

KEEHN, J. D. Beyond the law of effect. In, *Schedule-induced and schedule-dependent phenomena.* Vol. 1. Toronto: Addiction Research Foundation, 1970.

KIMBLE, G. A. *Hilgard and Marquis' conditioning and learning.* New York: Appleton-Century-Crofts, 1961.

LEVINE, M. Hypothesis behavior. In A. M. Schrier, H. F. Harlow, & F. Stollnitz (Eds.), *Behavior of nonhuman primates.* Vol. 1. New York: Academic Press, 1965.

LICKLIDER, J. C. R. On psychophysiological models. In W. A. Rosenblith (Ed.), *Sensory* *communication.* Cambridge: M.I.T. Press, 1961.

MacCorquodale, K. B. F. Skinner's *Verbal behavior:* A retrospective appreciation. *Journal of the Experimental Analysis of Behavior,* 1969, **12,** 831–841.

Mandler, G. The interruption of behavior. *Nebraska Symposium on Motivation,* 1964, **12,** 163–219.

Martin, I., & Levey, A. B. *The genesis of the classical conditioned response.* Oxford: Pergamon Press, 1969.

Mayr, E. The emergence of evolutionary novelties. In S. Tax (Ed.), *The evolution of life.* Vol. 1. Chicago: University Press, 1960.

Mayr, E. *Animal species and evolution.* Cambridge: Harvard University Press, 1963.

McFarland, D. J. Hunger, thirst and displacement pecking in the barbary dove. *Animal Behaviour,* 1965, **13,** 293–300.

McFarland, D. J. On the causal and functional significance of displacement activities. *Zeitschrift für Tierpsychologie,* 1966, **23,** 217–235.

McNeill, D. On theories of language acquisition. In T. R. Dixon & D. L. Horton (Eds.), *Verbal behavior and general behavior theory.* Englewood Cliffs, N.J.: Prentice-Hall, 1968.

Millenson, J. R., & Hurwitz, H. M. B. Some temporal and sequential properties of behavior during conditioning and extinction. *Journal of the Experimental Analysis of Behavior,* 1961, **4,** 97–106.

Miller, N. E. Learning of visceral and glandular responses. *Science,* 1969, **163,** 434–445.

Minsky, M. Steps toward artificial intelligence. *Proceedings of the Institute of Radio Engineers,* 1961, **49,** 10–30.

Morris, D. The reproductive behaviour of the Zebra finch (*Poephila guttata*), with special reference to pseudofemale behaviour and displacement activities. *Behaviour,* 1954, **6,** 271–322.

Morris, D. The reproductive behaviour of the Ten-spined Stickleback (*Pygosteus pungitius* L.). *Behaviour,* 1958, Supplement VI, 1–154.

Neimark, E. D., & Estes, W. K. *Stimulus sampling theory.* San Francisco: Holden-Day, 1967.

Neisser, U. *Cognitive psychology.* New York: Appleton-Century-Crofts, 1967.

Neuringer, A. J. Many responses per food reward with free food present. *Science,* 1970, **169,** 503–504. (a)

Neuringer, A. J. Superstitious key pecking after three peck-produced reinforcements. *Journal of the Experimental Analysis of Behavior,* 1970, **13,** 127–134. (b)

Pavlov, I. P. *Conditioned reflexes.* (Trans. by G. V. Anrep.) London: Oxford University Press, 1927.

Pfaffman, C. The behavioral science model. *American Psychologist,* 1970, **25,** 437–441.

Pringle, J. W. S. On the parallel between learning and evolution. *Behaviour,* 1951, **3,** 174–215.

Pryor, K. W., Haag, R., & O'Reilly, J. The creative porpoise: Training for novel behavior. *Journal of the Experimental Analysis of Behavior,* 1969, **12,** 653–661.

Rachlin, H. Autoshaping of key pecking in pigeons with negative reinforcement. *Journal of the Experimental Analysis of Behavior,* 1969, **12,** 521–531.

Rescorla, R. A. Pavlovian conditioning and its proper control procedures. *Psychological Review,* 1967, **74,** 71–80.

Revusky, S. H., & Bedarf, E. W. Association of illness with prior ingestion of novel foods. *Science,* 1967, **155,** 219–220.

Reynierse, J. H., & Spanier, D. Excessive drinking in rats' adaptation to the schedule of feeding. *Psychonomic Science,* 1968, **10,** 95–96.

Rosenblith, J. Z. Polydipsia induced in the rat by a second-order schedule. *Journal of the Experimental Analysis of Behavior,* 1970, **14,** 139–144.

Rowell, C. H. F. Displacement grooming in the Chaffinch. *Animal Behaviour,* 1961, **9,** 38–63.

Rozin, P. Central or peripheral mediation of learning with long CS–US intervals in the feeding system. *Journal of Comparative and Physiological Psychology,* 1969, **67,** 421–429.

Schaeffer, R. W., & Premack, D. Licking rates in infant albino rats. *Science,* 1961, **134,** 1980–1981.

Segal, E. The interaction of psychogenic polydipsia with wheel running in rats. *Psychonomic Science,* 1969, **14,** 141–144. (a)

Segal, E. Transformation of polydipsic drinking into operant drinking. A paradigm? *Psychonomic Science,* 1969, **16,** 133–135. (b)

Segal, E. Speculations on the provenance of operants. In, *Schedule-induced and schedule-dependent phenomena.* Vol. 2. Toronto: Addiction Research Foundation, 1970.

Segal, E., & Holloway, S. M. Timing behavior in rats with water drinking as a mediator. *Science,* 1963, **140,** 888–889.

Segal, E., Oden, D. L., & Deadwyler, S. A. Determinants of polydipsia: IV. Free-reinforcement schedules. *Psychonomic Science,* 1965, **3,** 11–12.

Sevenster, P. A causal analysis of a displacement activity (Fanning in *Gasterosteus Aculeatus* L.). *Behaviour,* 1961, Supplement IX, 1–170.

Shanab, M. E., & Peterson, J. L. Polydipsia in the pigeon. *Psychonomic Science,* 1969, **15,** 51–52.

Sheffield, F. S. Relation between classical conditioning and instrumental learning. In W. F. Prokasy (Ed.), *Classical conditioning: A symposium.* New York: Appleton-Century-Crofts, 1965.

Shimp, C. P. Optimal behavior in free-operant experiments. *Psychological Review,* 1969, **76,** 97–112.

Sidman, M., & Fletcher, F. G. A demonstration of auto-shaping with monkeys. *Journal of the*

Experimental Analysis of Behavior, 1968, **11**, 307–309.

Simmelhag, V. L. The form and distribution of responding in pigeons on response-independent fixed and variable interval schedules of reinforcement. Unpublished master's thesis, University of Toronto, 1968.

Simpson, G. G. *The major features of evolution.* New York: Columbia University Press, 1953.

Skinner, B. F. *The behavior of organisms.* New York: Appleton-Century, 1938.

Skinner, B. F. "Superstition" in the pigeon. *Journal of Experimental Psychology,* 1948, **38**, 168–172.

Skinner, B. F. *Science and human behavior.* New York: Macmillan, 1953.

Skinner, B. F. Reinforcement today. *American Psychologist,* 1958, **13**, 94–99.

Skinner, B. F. An experimental analysis of certain emotions. *Journal of the Experimental Analysis of Behavior,* 1959, **2**, 264.

Skinner, B. F. *Cumulative record.* New York: Appleton-Century-Crofts, 1961.

Skinner, B. F. Operant behavior. In W. K. Honig (Ed.), *Operant behavior: Areas of research and application.* New York: Appleton-Century-Crofts, 1966. (a)

Skinner, B. F. The phylogeny and ontogeny of behavior. *Science,* 1966, **153**, 1205–1213. (b)

Skinner, B. F. *Contingencies of reinforcement.* New York: Appleton-Century-Crofts, 1969.

Skinner, B. F., & Morse, W. H. Fixed-interval reinforcement of running in a wheel. *Journal of the Experimental Analysis of Behavior,* 1958, **1**, 371–379.

Staddon, J. E. R. Some properties of spaced responding in pigeons. *Journal of the Experimental Analysis of Behavior,* 1965, **8**, 19–27.

Staddon, J. E. R. Asymptotic behavior: The concept of the operant. *Psychological Review,* 1967, **74**, 377–391.

Staddon, J. E. R. Inhibition and the operant. A review of G. v. Békésy, *Sensory inhibition,* and F. Ratliff, *Mach bands: Quantitative studies on neural networks in the retina. Journal of the Experimental Analysis of Behavior,* 1969, **12**, 481–487.

Staddon, J. E. R. Reinforcement after-effects. In, *Schedule-induced and schedule-dependent phenomena.* Vol. 2. Toronto: Addiction Research Foundation, 1970. (a)

Staddon, J. E. R. Temporal effects of reinforcement: A negative "frustration" effect. *Learning and Motivation,* 1970, **1**, 227–247. (b)

Stein, L. Excessive drinking in the rat: Superstition or thirst? *Journal of Comparative and Physiological Psychology,* 1964, **58**, 237–242.

Stellar, E., & Hill, J. H. The rat's rate of drinking as a function of water deprivation. *Journal of Comparative and Physiological Psychology,* 1952, **45**, 96–102.

Stricker, E. M., & Adair, E. R. Body fluid balance, taste, and postprandial factors in schedule-induced polydipsia. *Journal of Comparative and Physiological Psychology,* 1966, **62**, 449–454.

Terrace, H. S. Stimulus control. In W. K. Honig (Ed.), *Operant behavior: Areas of research and application.* New York: Appleton-Century-Crofts, 1966.

Thompson, D. M. Escape from S^D associated with fixed-ratio reinforcement. *Journal of the Experimental Analysis of Behavior,* 1964, **7**, 1–8.

Tinbergen, N. "Derived" activities: Their causation, biological significance, origin, and emancipation during evolution. *Quarterly Review of Biology,* 1952, **27**, 1–32.

Tulving, E., & Madigan, S. A. Memory and verbal learning. *Annual Review of Psychology,* 1970, **21**, 437–484.

Upton, M. The auditory sensitivity of guinea pigs. *American Journal of Psychology,* 1929, **41**, 412–421.

Valenstein, E. S., Cox, V. C., & Kakolewski, J. W. Reexamination of the role of the hypothalamus in motivation. *Psychological Review,* 1970, **77**, 16–31.

Verplanck, W. S. Since learned behavior is innate, and vice versa, what now? *Psychological Review,* 1955, **52**, 139–144.

von Holst, E., & von Saint Paul, U. On the functional organisation of drives. *Animal Behaviour,* 1963, **11**, 1–20.

Williams, D. R. Classical conditioning and incentive motivation. In W. F. Prokasy (Ed.), *Classical conditioning: A symposium.* New York: Appleton-Century-Crofts, 1965.

Williams, D. R., & Williams, H. Auto-maintenance in the pigeon: Sustained pecking despite contingent non-reinforcement. *Journal of the Experimental Analysis of Behavior,* 1969, **12**, 511–520.

Wolin, B. R. Difference in manner of pecking a key between pigeons reinforced with food and with water. In A. C. Catania (Ed.), *Contemporary research in operant behavior.* New York: Scott, Foresman, 1968.

Wright, S. Evolution in Mendelian populations. *Genetics,* 1931, **16**, 97–159.

Wüttke, W. The effects of d-amphetamine on schedule-controlled water licking in the squirrel-monkey. *Psychopharmacologia* (Berlin), 1970, **17**, 70–82.

Zeaman, D., & Smith, R. W. Review of some recent findings in human cardiac conditioning. In W. F. Prokasy (Ed.), *Classical conditioning: A symposium.* New York: Appleton-Century-Crofts, 1965.

Zener, K. The significance of behavior accompanying conditioned salivary secretion for theories of the conditioned response. *American Journal of Psychology,* 1937, **50**, 384–403.

(Received March 19, 1970)

Psychological Review
1971, Vol. 78, No. 1, 44–57

THE DRIVE THEORY OF SOCIAL FACILITATION[1]

ROBERT FRANK WEISS[2] AND FRANKLIN G. MILLER
University of Oklahoma

The scope and power of the drive theory of social facilitation are extended by taking fuller advantage of learning theory models of drive. Existing theory employs the irrelevant drive paradigm, in which neither the initiation nor the termination of audience observation is contingent on the behavior of the subject or on the onset of the conditioned stimulus. Extensions include escape and avoidance of audience observation, as well as the classical conditioning of symbolic audiences. In addition, nine methods for varying the strength of audience-induced drive are developed, based on the model of learned drive, including extinction, summation, generalization, acquisition, and five forms of inhibition.

The effects of the presence of others on the behavior of an individual is generally recognized as one of the traditional central problems of social psychology. The study of such effects has come to be called the problem of social facilitation, despite the fact that the presence of others sometimes facilitates and sometimes impairs performance. In a theoretical essay of great elegance, Zajonc (1965) reconciled the apparently contradictory results of numerous experiments on social facilitation. Zajonc proposed that the presence of others, whether audience or coactors, serves to arouse drive (*D*). Since generalized drive energizes all habits, both correct and incorrect responses will be facilitated, with the greatest benefits going to the best learned dominant responses (e.g., Spence, 1956). Cottrell (1968) greatly sharpened this theoretical analysis by (*a*) replacing the overinclusive concept of dominant responses with the more correct learning theory concept of stronger habit strength (*H*), (*b*) identifying and demonstrating the learned drive basis of *D*, (*c*) clarifying the boundary conditions of the effect, and (*d*) emphasizing the importance of using experimental tasks which are already known to reflect the effects of drive. It now seems quite clear that the drive induced by audience observation is not an appetitive drive like hunger or thirst, but an aversive one like frustration or anxiety. This point follows directly from Cottrell's learned drive interpretation, since the evidence is abundant that learned drives are based on noxious primary drives like pain (e.g., Miller, 1948) and frustration (e.g., Daly, 1969; Wagner, 1963) and cannot be based on appetitive drives like hunger and thirst (e.g., Cravens & Renner, 1969; Myers & Miller, 1954; Novin & Miller, 1962). Social psychological experiments on the effects of audience observation on nervous sweating (Martens, 1969a, 1969b) also support this conclusion.[3]

A rapidly expanding research literature has been directly inspired by the drive theory of social facilitation (e.g., Cottrell, Rittle, & Wack, 1967; Henchy & Glass, 1968; Klinger, 1969; Martens, 1969a; Matlin & Zajonc, 1968; Weiss, Miller, Langan, & Cecil, 1971). In these studies, the socially facilitated subject typically works on a learning task while an audience observes his

---

[1] This research was supported in part by Grant MH-12402 from the National Institute of Mental Health to the senior author.

[2] Requests for reprints should be sent to Robert Frank Weiss, Department of Psychology, University of Oklahoma, Norman, Oklahoma 73069.

[3] Among the particular drive mechanisms for social facilitation which have been suggested are fear and anticipatory frustration (Cottrell, 1968; Lott & Lott, 1968) as well as conflict-induced drive (Thibaut & Kelley, 1959). Variations on fear drive might include shame, embarrassment, effectance, and, occasionally, guilt. Cottrell (1968) and Lott and Lott (1968) briefly entertain the possibility of facilitation based on incentive motivation, but the available evidence does not support a Spencian hypothesis of drive-like energization of preexisting responses by conditioned incentive (Armus, Carlson, Guinan, & Crowell, 1964; Armus & Sniadowski-Dolinsky, 1966; Trapold, 1962; Trapold & Winokur, 1967).

performance. The presence of the audience induces irrelevant drive (D), which energizes all responses in the situation. The habit strength (H) of the various responses may be manipulated with the result that the greatest benefits of differences in drive go to the strongest habits, whether these are "correct" or "incorrect." Thus, all the studies based on the drive theory of social facilitation have used social analogs of the standard learning paradigm of a test for the energizing effect of an *irrelevant* drive on behavior, in which neither the onset nor the reduction of drive is contingent on the behavior of the subject or the presentation of the CS. However, learning theory ascribes other effects to drive besides response energization, which may be studied in paradigms in which drive is relevant to the stimulus–response–reinforcement contingencies. Thus, the reduction of noxious drives is responsible for conditioning instrumental escape and avoidance responses. In other paradigms, drive stimuli serve a punishing function. Learning theory and its related research also offer a fairly detailed statement of the manner in which the contingency between a CS and drive onset leads to the conditioning of learned drives. The purpose of this paper is to extend and deepen the drive theory of social facilitation by specifying the implications of learning models for predicting the effects of the presence of others on behavior, beyond the confines of the irrelevant drive paradigm. The paradigms to be discussed include escape and avoidance conditioning, punishment, and the classical conditioning of learned drives. For each paradigm we shall summarize the learning model, derive hypotheses concerning audience effects on performance, and suggest some ways of implementing the empirical tests.

Before discussing the four paradigms just mentioned, it will be useful to make more explicit the theory construction methods employed, and then to specify the methods for manipulating the strength of audience-induced drive which are essential to the discussion of all four paradigms.

METHOD

In this paper, the general approach which Miller (1959) has called "extension of liberalized S–R theory" is employed to further develop the Zajonc-Cottrell drive theory of social facilitation. Hull-Miller-Spence learning theory and its related research (referred to frequently hereafter simply as "learning theory" or "learning principles") (e.g., Amsel, 1967; Brown & Farber, 1968; D'Amato, 1969; Dollard & Miller, 1950; Hull, 1943, 1952; Logan, 1959; Spence & Spence, 1966; Spence, 1956, 1966) is used as a model for social facilitation paradigms corresponding to instrumental escape conditioning, avoidance conditioning, punishment, and classical conditioning of learned drives.

Hullian learning theory has been developed primarily to predict individual behavior in highly controlled experimental situations. Nevertheless, learning theory has been extended not only to social facilitation but to numerous social processes with a considerable measure of success (e.g., Lott & Lott, 1968; Miller & Dollard, 1941; R. F. Weiss, 1963, 1968; Wheeler, 1966; Whiting & Child, 1953; Wolpe, 1958). The explanatory power of learning theory stems, in part, from three sources. First, learning theory principles cover a variety of basic learning paradigms, such as classical defense conditioning, instrumental escape conditioning, etc. Second, learning theory principles may be combined in a determinate manner, becoming more powerful explanatory tools when the manner of their interaction is specified. Third, learning theory is quantitative, with the usual advantages that attend scientific quantification.

Technique of Theory Construction

The use of a model in theory construction typically involves the specification of a dictionary of analogies (rules of correspondence) which relate the variables of the model to the variables of the research area to be explained and predicted. Once this is done, the relations holding among the variables of the model must, theoretically, also hold between the corresponding (analogous) variables in the data area to be explained (e.g., Brodbeck, 1959; Campbell, 1920; Lachman, 1960; Oppenheimer, 1956). The systematic use of learning theory as a model for social facilitation makes it possible to take full advantage of the three characteristics of learning theory mentioned above: principles covering a variety of learning paradigms, combination of principles in a determinate manner, and quantitative specification. Thus, for example, social facilitation research has never taken advantage of the numerous learning principles applicable to an escape conditioning paradigm. Three such principles regarding the effects of delay of reinforcement on measures of escape conditioning may be stated informally in the following manner: (a) the delay of reinforcement gradient is decreasing and negatively accelerated in shape;

(*b*) delay of reinforcement and number of trials combine multiplicatively; and (*c*) shorter delays of reinforcement maximize the effects of differences in magnitude of reinforcement. If a social variable is to be analogous to delay of reinforcement in escape conditioning, its effects on appropriate dependent measures must show that (*a*) this social variable gradient is decreasing and negatively accelerated in shape, (*b*) this social variable and number of trials combine multiplicatively, and (*c*) smaller values of this social variable maximize the effects of differences in magnitude of reinforcement.

The above points illustrate the development of an analogy between learning and social *independent* variables. The analogy does not have testable implications until analogies between learning and social *dependent* variables are also developed. Moreover, it is necessary to clearly specify *which learning paradigm* the social conditions are analogous to; instrumental reward conditioning, instrumental escape conditioning, selective learning, etc., differ sharply in certain regards.

Models and Translations

The theories developed here are not mere translations of one behavioral science language into another, since the theories generate hypotheses which are not in the existing literature. A "mere translation" would require, for example, that the hypotheses concerning the previous paragraph's social variable already exist in the literature of social facilitation research before they could be merely translated into statements about delay of reinforcement and number of trials.

Analogy and Reduction

Nowhere in this essay is it proposed that social behavior represents a distinct higher level of phenomena which can only be fully explained by reducing it to a lower level of phenomena such as conditioning. Nor is this dubious philosophical program implicit in our intentions. Rather, the intention has been to use the relatively well understood as a model for that which is presently less well understood. The programmatic assumptions involved are commonplace to scientific method: Well-constructed models and analogies help to stimulate and guide research and to integrate broad ranges of knowledge through an underlying set of common principles.

Nature of Reinforcement

The theories developed here are not specifically concerned with the ultimate nature of reinforcement: an empirical law-of-effect position is taken throughout. The empirical law of effect recognizes the fact that reduction of noxious stimulation following a response is reinforcing, whereas the more controversial drive reduction theory of reinforcement holds, in its fully developed form, that all reinforcement is based on drive reduction. The drive reduction theory is not essential to Hull-Miller-Spence learning theory. Hullian theory predicts and explains in the same manner without reference to this particular conception of the nature of reinforcement (Dollard & Miller, 1950; Miller, 1959; Spence, 1956).

MANIPULATING DRIVE STRENGTH

In the discussion of the escape, avoidance, punishment, and classical conditioning paradigms for social facilitation which follow the present section, variation of the strength of audience-induced drive occupies a prominent place, and it will be useful, therefore, to first explore methods for manipulating drive strength. If the drive induced by audience observation is a learned one, as theory and experiment indicate (e.g., Cottrell, 1968; Cottrell, Wack, Sekerak, & Rittle, 1968; Henchy & Glass, 1968; Klinger, 1969; Lott & Lott, 1968; Paulus & Murdoch, in press), it should be possible to extend some of the familiar techniques for the manipulation of learned drive strength to the social facilitation situation. These techniques include extinction, inhibition, summation, generalization, and acquisition.

Extinction

The strength of a learned drive may be decreased by repeated presentation of the CS which arouses that drive, unaccompanied by the aversive UCS (e.g., Armus, 1960; Kalish, 1954). If audience observation arouses a learned drive, it should be possible to decrease the drive-arousing power of audience observation by repeated exposures to that audience without noxious consequences following from that observation.

Inhibition

Since learned drives manifest a high resistance to extinction, a more powerful method of decreasing the strength of audience-induced drive may be to elicit a response which is incompatible with the learned drive. One version of this method would be to give subjects reciprocal inhibition therapy (Wolpe, 1958) for "audience phobia" (Paul, 1967). Another version would be to countercondition against the learned fear by repeated presentations of audience observation followed by pleasant rewarding consequences (Melvin & Brown, 1964). Even without formal counterconditioning trials, making the goal response to a primary reward object may be drive reducing (Wilson & Dinsmoor, 1970; but see also Nelson, 1966), suggesting the applicability of Razran's (1938) luncheon technique to social facilitation research. Similarly, there is some evidence that a signal which predicts positive reward will inhibit fear drive (Bull & Overmier, 1969); a signal which indicates that pleasant rewarding consequences will follow audience observation should then serve as an inhibitor of audience-induced drive. Indeed, a signal which predicts mere safety from shock will inhibit a learned fear drive (e.g., Bull & Overmier, 1968; Rescorla, 1969; Rescorla & LoLordo, 1965); a signal which indicates that negative consequences will not follow audience observation should then serve as an inhibitor of audience-induced drive. Such a signal could, of course, be verbal.

Summation

If a learned drive is independently conditioned to more than one CS, the strength of the learned drive may be increased by presenting these conditioned stimuli simultaneously (McAllister & McAllister, 1965, 1967). If a learned drive is aroused by observers, it may be possible to vary the strength of the learned drive by varying the number of observers in the audience.

Generalization

Learned drives may be weakened by reducing the similarity of the drive-arousing CS to the CS originally used in establishing the learned drive. Similarly, learned drives will generalize to a new situation to the extent that the new drive-arousing CS resembles the original one (e.g., Desiderato, 1964). Thus, it should be possible to vary the strength of the learned drive aroused by audience observation by varying the similarity of the audience involved to those audiences whose observation and evaluation have typically been followed by harsh consequences. The important finding of Cottrell et al. (1968) that blindfolded observers do not arouse enough drive to produce a social facilitation effect may be regarded as an extreme case of this generalization mechanism. Less drastic variations might include the adoption of different degrees of evaluative stance by the observers (Paulus & Murdoch, in press), or the observation of subjects by observers who typically exercise differential hierarchal or peer-level punitive power over them—as the observation of white-collar workers by (*a*) managers, or (*b*) other white-collar workers, or (*c*) blue-collar workers—or the observation of college students by other students or by children. Given sufficient information on the subjects, the possibilities of arousal of the shame and guilt forms of learned fear (e.g., Aronfreed, 1968; Dollard & Miller, 1950) are intriguing. Consider the possibilities of facing subjects who have stronger feelings of guilt and shame toward their mothers than toward their fathers with audiences consisting of women or men, or the reactions of (measurably) guilty whites to white or black observers.

Acquisition

Finally there is the possibility of trying to directly condition learned drives to the audience in order to increase its ability to arouse drive, for example, by following audience observation with harsh evaluation or other noxious stimuli, prior to employing this audience in a typical social facilitation experiment. The acquisition method permits a laboratory simulation of the natural origins (Cottrell, 1968) of audience effects. It may well be possible to condition drive to a particular audience, without requiring the audience to judge the subject harshly, by means of higher order conditioning (e.g., Anderson, Johnson, & Kempton, 1969; Kamil, 1969; McAllister & McAllister, 1964). Aronfreed (1968) argues that learned shame is aroused by highly visible incompetence. If the experimenter arranges for the subject to perform incompetently before an audience, shame should be elicited, and a higher order learned drive of shame should then become conditioned to that particular audience.

Escape Conditioning

Primary Concepts

One of the most important tools of thought provided by learning theory is the Thorndikian tradition of the law of effect (e.g., Berger & Lambert, 1968; Skinner, 1963).

In this instrumental learning tradition, attention is focused on the contingency between a response and its reinforcing consequences, rather than on the stimulus antecedents of a response. In the typical irrelevant drive paradigm for audience effects, we focus on the stimulus antecedents of a response: The subject is exposed to the audience stimulation and the effects on behavior are observed. In the escape conditioning paradigm, we avail ourselves of the opportunity of the Thorndikian tradition to focus on the contingency between a response and its reinforcing consequences. In a standard learning experiment on escape conditioning, reduction of an aversive drive, such as electric shock, is made contingent on the performance of an instrumental response, such as bar pressing. The reduction of the aversive drive serves to reinforce the instrumental response. In social facilitation research, the aversive drive is provided by audience observation, and the termination of this observation should reinforce a response which precedes it. A concrete example of a possible social facilitation escape conditioning experiment, phrased in Miller and Dollard's (1941) familiar terms of drive, cue, response, and reinforcement, follows. The deception procedures needed in a finished experiment are omitted here, and leave considerable room for experimenter ingenuity. Two observers watch and evaluate a subject through a transparent glass screen and microphones. They are clearly visible to the subject and serve to induce drive. The social drive-reducing instrumental response to be learned is button pushing. Pushing the button opaques the screen and turns off the microphone, terminating the observation for some preset period of time, thus reinforcing the subjects. The cue for responding is the removal of a cover from the button. The time between this cue and the social drive-reducing instrumental response is response latency, and latency, or its reciprocal, response speed, is the dependent variable. If audience observation induces an aversive drive, the subjects should learn to press a button in order to terminate this observation.

Independent Variables: Rules of Correspondence

The possibilities of the escape conditioning model are not exhausted by the single informal prediction above. If rules of correspondence between escape conditioning variables and social facilitation variables are specified, then the relations holding among the escape conditioning variables should, in theory, hold among the corresponding social facilitation variables. The rules of correspondence relating the variables of escape conditioning to the variables of social facilitation are given below, and numbered for later reference.[4] Several kinds of trials are possible in this paradigm. Corresponding to a reinforced trial in escape conditioning is a social facilitation trial in which the audience observation begins before the social drive-reducing instrumental response and is terminated by it (1). Corresponding to a nonreinforced trial in escape conditioning is a social facilitation trial in which the audience observation begins before the response, but the response does not terminate the aversive observation (2). A series of such nonreinforced trials following acquisition is analogous to extinction (3), while a mixture of trials on which the button push sometimes terminates the audience observation and sometimes does not is analogous to partial reinforcement (4). It is also possible to have a trial which is not observed by an audience, corresponding to a procedure sometimes used in escape conditioning in which the shock is altogether omitted from some trials (5). A mixture of such unobserved trials with the normal reinforced social facilitation trials is analogous to the intermittent shock schedules introduced by Jones (1953) and recently studied by Franchina (e.g., 1966, 1969a) (6).

---

[4] In a philosophical exercise designed to maximize parsimony, it could easily be shown that these rules of correspondence could be replaced by a small set of fundamental rules and a larger set of rules which are logical corrollaries of the fundamental rules. Similarly, in the interests of clarity and simplicity of exposition, the theory is stated solely in terms of audience effects, but is, of course, intended to be applicable to coaction effects as well.

The intensity of drive induced through audience observation can be manipulated by the several techniques described earlier (7). In escape conditioning the shock is typically reduced to a zero level, following a response, on a reinforced trial. When it is necessary to vary magnitude of reinforcement independently of the intensity of drive, this may be done by reducing the shock by large or small amounts following the response. If drive intensity depends at all on the number of observers, it might be possible to vary an analog of magnitude of reinforcement by reducing the number of observers by large or small amounts following the response (8). A procedure precisely analogous to this has been successfully employed in fear-motivated escape conditioning by McAllister and McAllister (1967), who motivated their subjects by two simultaneously presented conditioned fear stimuli and reinforced them by terminating one fear stimulus for the low-reinforcement group and both fear stimuli for the high-reinforcement group. An interesting variation of this procedure might be an audience consisting of representatives of various social sectors, such as blue-collar and white-collar observers, negro and white observers, or male and female observers, with reinforcement terminating the observation of one or both social sectors. Finally, the time interval between the instrumental response and the termination of observation corresponds to delay of reinforcement (9).

Dependent Variables: Rules of Correspondence

In the preceding sections we have specified the kind of learning paradigm to which our social facilitation paradigm is analogous, and we have developed a dictionary of analogies between learning and social facilitation independent variables. The failure to use appropriate dependent variables is one of the most common failings of social learning research. Cottrell (1968) and Weiss (1962, 1968) have strongly emphasized the importance of choosing dependent variables which are appropriate to the theory to be tested. Indeed, "the theory does not have testable implications until analogies between learning and social *dependent* variables are developed [R. F. Weiss, 1968, p. 111]." In Hullian learning theory, speed (1/latency) and resistance to extinction are theoretically appropriate dependent measures of escape conditioning, although the extinction measure is not very well researched. Latency corresponds quite straightforwardly to the time between the cue and the button-push response in the above example (10). Latency measures have been used successfully in social facilitation research, in the usual irrelevant drive paradigm (Matlin & Zajonc, 1968; Zajonc, Heingartner, & Herman, 1969). Changes in speed during trials on which responding no longer terminates or decreases observation may be used as measures of resistance to extinction, if corrected for differences in acquisition asymptote (Anderson, 1963) (11). Alternatively, such trials may be given until the subject ceases to respond, according to some explicit criterion of cessation, and the number of trials until this criterion is met may be counted (12).

Predictions

Rules of correspondence have been assumed between the variables of escape conditioning and the variables of social facilitation. Once such rules of correspondence are assumed, the functional relationships holding among the variables of the conditioning model should also hold among the corresponding social facilitation variables. The following nine derivations are illustrative, not exhaustive.

Acquisition. In escape conditioning, response speed is an increasing function of the number of reinforced trials. Given Rules of Correspondence 1 and 10, it therefore follows that speed of the social drive-reducing instrumental response should be an increasing function of the number of trials on which audience observation begins before the button push and is terminated by it.

Extinction. In escape conditioning, response speed is a decreasing function of the number of unreinforced trials. Given Rules of Correspondence 3 and 10, it therefore follows that speed of the social drive-reducing instrumental response should be a decreasing function of the number of audience-observed trials on which the button push does not terminate the observation.

Partial reinforcement. In escape conditioning, speed is slower under partial reinforcement than it is under continuous reinforcement (Bower, 1960). Given Rules of Correspondence 4 and 10, it therefore follows that subjects given a mixture of trials on which the response sometimes terminates, the audience observation will respond more slowly than subjects whose response always terminates the observation. It is important to note that the effects of partial reinforcement on asymptotic performance in escape conditioning are the reverse of those found in instrumental reward conditioning, and serve as one of the several ways of testing the difference between an aversive and an appetitive drive.

Shock schedules. In escape conditioning, response speed is a decreasing function of the percentage of trials on which shock is omitted (Franchina, 1966, 1969a). Given Rules of Correspondence 6 and 10, it therefore follows that the speed of the social drive-reducing instrumental response is a decreasing function of the percentage of unobserved trials.

Magnitude of reinforcement. In escape conditioning, response speed is an increasing function of the magnitude of reinforcement (Bower, Fowler, & Trapold, 1959; McAllister & McAllister, 1967). Given Rules of Correspondence 8 and 10, t therefore follows that the speed of the ocial drive-reducing instrumental response hould be an increasing function of the reuction in the number of observers or oberver classes.

Drive intensity. Studies controlling the agnitude of reinforcement have yielded nflicting results (Campbell & Kraeling, 53; McAllister & McAllister, 1967; oods, Davidson, & Peters, 1964). However, when shock is terminated completely er each escape, there is no doubt that eed is an increasing function of drive g., Franchina, 1969b; Trapold & Fowler, 60). Given Rules of Correspondence 7 d 10, it therefore follows that speed of an strumental response which terminates audice observation should be an increasing function of the several drive-increasing operations discussed near the beginning of the paper.

Delay of reinforcement. In escape conditioning, response speed is a negatively accelerated decreasing function of delay of reinforcement (Fowler & Trapold, 1962). Given Rules of Correspondence 9 and 10, it therefore follows that speed of the social drive-reducing instrumental response should be a negatively accelerated decreasing function of the delay in termination of observation.

Interaction: Delay × Trials. In escape conditioning, delay of reinforcement and number of reinforced trials combine multiplicatively to determine response speed (Fowler & Trapold, 1962). Given Rules of Correspondence 9, 1, and 10, it therefore follows that the delay in termination of observation and the number of trials on which observation is terminated following the response should combine multiplicatively to determine the speed of the social drive-reducing instrumental response. Delay effects should thus be greatest late in training.

Interaction: Delay × Magnitude. In escape conditioning, magnitude of reinforcement and delay of reinforcement interact to determine response speed (Woods & Feldman, 1966). Magnitude effects are greatest at short delays, and delay effects greatest at high magnitudes. Given Rules of Correspondence 9, 8, and 10, it therefore follows that the reduction in the number of observers or observer classes and the delay in this reduction should combine in the above manner to determine the speed of the instrumental response. If audience observation induced an appetitive drive, a quite different outcome would be expected (Logan, 1960).

Boundary Conditions

Any theory is intended to apply only within a limited range of conditions.

> Ideally all such limitations should be stated explicitly by the theorist, although, in practice, he may not have recognized some of these and may not yet be ready to specify others [Logan, 1959, p. 300].

Since the theory of social facilitation developed here is based on analogies with conditioning research and theory, some of the most essential boundary conditions are also based on analogies with conditioning.

Conditioning versus response competition. In conditioning, the aim of the experimenter is to "limit the investigation to the study of the strengthening of a single response" and the experimenter therefore attempts "to minimize the occurrence of competing responses [Spence, 1956, p. 38]." These strictures apply directly to the conditioning of the socially facilitated instrumental response. The differences between drive effects in the single response and competing response situations are well known to social facilitation researchers (e.g., Cottrell et al., 1967; Zajonc et al., 1969).

Trials and dependent variables. Following Zajonc (1965) and Cottrell (1968), the analysis presented here is based on the Hullian theory of drive (e.g., Spence, 1956). As previously noted, appropriate dependent variables for testing this theory in the instrumental escape conditioning paradigm are the speed (1/latency) and resistance to extinction of the social drive-reducing instrumental response. These measures are obtained in a discrete-trials paradigm such as the previously described example. It is not possible, therefore, to test the theory presented here by means of a free-responding operant type of instrumental learning task in which response rate rather than response latency is the dependent variable. While the discrete-trials and free-responding operant escape paradigms differ from each other in essential ways, most of the rules of correspondence for independent variables can be adapted to a free-responding paradigm with relatively little difficulty. The investigator who prefers such a free-responding paradigm should note, however, the specific properties of the free-responding escape model (Dinsmoor, 1968) and recall that this paradigm is irrelevant to Hullian theories of drive.

In Hullian theory, the amplitude, (i.e., vigor or forcefulness of a response) has regularly been used as a measure of the strength of classically conditioned autonomic responses (e.g., Hull, 1943). It is worth explicitly mentioning, therefore, that such measures do not provide an index of the strength of a response in instrumental conditioning of motor responses, whether in the discrete-trials or free-responding paradigm. For example, the amplitude of an instrumental response does not increase over the course of acquisition (e.g., Notterman & Mintz, 1965; Weiss, 1961).

Drive arousal. The theoretical predictions apply only to the case in which the presence of an audience increases drive. Cottrell et al. (1968) have made it crystal clear that there are specifiable conditions under which the presence of an audience will not induce drive. This boundary applies generally to the drive theory of social facilitation, rather than being special to the present extensions of it, and it seems appropriate, therefore, to keep the treatment of the topic in this paper fairly brief. If the drive induced by audience observation is a learned one, then it follows that there are two general conditions under which the presence of an audience would be sufficient to increase drive, at least one of which would be necessary to increase drive. First, the audience (including audience behavior) is a conditioned stimulus for a learned drive of fear or anticipatory frustration (Cottrell, 1968; Lott & Lott, 1968). Variations on fear drive might include shame, embarrassment, effectance, and, occasionally, guilt. While not addressed, in so many words, to the problem of social facilitation, the accounts of the social conditions for the learning or arousal of such drives in Dollard and Miller (1950, especially pp. 116 ff., 243–245) as well as in Aronfreed (1968, pp. 241 ff.) and Byrne (1969) are indispensible for social facilitation research. An evaluative stance on the part of the audience is the most important factor in fear arousal so far isolated in the social facilitation situation itself (Cottrell, 1968; Henchy & Glass, 1968; Klinger, 1969; Paulus & Murdoch, in press). Variations on the conditioned stimulus function of the audience include the possibility of audience arousal of a cue-producing mediating response ("It's a test") which then serves as the immediate CS for the learned drive, or the possibility of the audience serving as only one part of a compound cue to which the learned drive is conditioned (so that evaluative observation of a manual versus an intellectual task might prove differentially drive arousing). A second (and secondary) way in which the audience can increase drive is by weakening the effect of an inhibitor of drive. Thus, for example, a child in a stressful situation may inhibit anxiety by sucking his thumb, but he can be deterred from sucking his thumb by the presence of an audience. A second example is of wider significance. Fear is inhibited by subject control of a frightening situation (e.g., Mowrer & Viek, 1948; J. M. Weiss, 1968) and the presence of an evaluative audience may have the effect of shifting some of the control of the situation from the subject to the evaluative audience.

The effects of the presence of others on drive are not limited to drive increase when the above conditions are met and no increase when these conditions are not met. There are, of course, theoretical reasons for thinking that under some circumstances the presence of others can reduce drive (e.g., Bovard, 1959; Byrne, 1969; Dollard & Miller, 1950), and some evidence to support this is available from both human (Amoroso & Walters, 1969; Kiesler, 1966) and animal (e.g., Davitz & Mason, 1955; Morrison & Thatcher, 1969) studies. The conditions determining such effects are being elucidated under such rubrics as interpersonal attraction (e.g., Byrne, 1969; Lott & Lott, 1968) and animal gregariousness (e.g., Morrison & Thatcher, 1969) as well as social facilitation. While recognizing the relevance of attraction to social facilitation, it is not suggested that the two processes are simply congruent. Thus, for example, while the presence of people to whom we are attracted may normally be drive reducing, an evaluative audience of such people may be particularly drive arousing, especially when there is a realistic probability of a deleterious effect on a continuing association.

Drive reduction. Termination of audience observation should reduce drive in a manner analogous to termination of shock in escape conditioning. Simple and precise control of such drive reduction leaves much to be desired, even in the conditioning laboratory. Thus, in escape conditioning, termination of the shock does not actually reduce drive to zero, in part because of perseverative physiological effects of prior stress, but also in part because fear (anticipatory of renewed shock) becomes conditioned to the static cues of the situation (McAllister & McAllister, 1965). Similarly, in classical defense conditioning of human subjects, it is believed that termination of the noxious airpuff does not actually reduce drive to zero because fear (r_e, anticipatory of renewed noxious stimulation) becomes conditioned to the cues of the situation (e.g., Spence, Haggard, & Ross, 1958; Trapold & Spence, 1960). An analogous difficulty can safely be expected in social facilitation research, where fear, anticipatory of renewed audience observation, would prevent termination of audience observation from reducing drive to zero. Providing that the main and interactive effects of such anticipations are not substantially greater than in escape conditioning, social facilitation data should prove to resemble escape conditioning data.

Deception. The deception procedures which social psychologists have mastered so skillfully are necessary for the study of the conditioning of voluntary responses in articulate organisms (e.g., Spence, 1966).

AVOIDANCE AND PUNISHMENT

If subjects will learn to escape audience observation, they should also learn to avoid it, if they are given the opportunity. This opportunity may be given by presenting the cue for responding before the onset of the aversive audience observation, and allowing the instrumental response to prevent the onset of audience observation. Most of the rules of correspondence relating the independent variables of social facilitation to the independent variables of escape conditioning can be adapted to the avoidance paradigm with little difficulty.

The punishment paradigm requires that the subject has already learned an instrumental response. The situation may then be arranged so that the instrumental response is followed by the onset of the audience observation. If audience observation is aversive, this procedure should lead to the suppression of the instrumental response rather than to its facilitation. Given this basic paradigm, specific rules of correspondence are suggested rather directly by the previous sections. The properties of the punishment model are not identical with those of the avoidance and escape models. Thus, for example, learning research leads us to expect that (*a*) the amount of social suppression should be an increasing function of the strength of the audience-induced drive (e.g., Appel & Peterson, 1965), (*b*) at least when the number of punished trials is low, the amount of social suppression should be a decreasing function of the delay between the instrumental response and the onset of audience observation (e.g., Kamin, 1959), and (*c*) a well-learned instrumental response should be *less* resistant to social suppression than one which has received relatively few rewarded trials (Karsh, 1962; Miller, 1960).

CLASSICAL CONDITIONING

If a previously neutral CS regularly precedes the onset of an aversive drive, that CS will acquire the capacity to elicit a learned drive. This process is usually described in terms of classical conditioning. If a previously neutral CS regularly precedes the onset of audience observation, that CS should acquire the capacity to elicit a higher order learned drive. Thus, a previously neutral cue comes to serve as a symbol for audience observation. Since response-produced cues can serve in this capacity, it can scarcely be doubted that the acquisition of such symbolic audiences plays a role in the development of internalized shame. In a broader context, we see that this form of social facilitation research affords us a perspective on a classic problem of social psychology, which George Herbert Mead (1934) referred to as "bringing the social process into individual experience." The symbolic audience is, albeit imperfectly, related to the symbolic "other" of Mead.[5]

Rules of Correspondence

By analogy with the classical conditioning of learned drives, there are a number of independent variables which should affect the strength of the symbolic audience drive. The number of classical conditioning trials corresponds to the number of times the CS precedes audience observation. A nonreinforced trial corresponds to a trial on which the CS is presented but not followed by audience observation; extinction and partial reinforcement analogies follow obviously. The procedure in which the UCS is presented unpaired with the CS corresponds to a procedure in which audience observation occurs without being preceded by the CS. The CS-UCS interval corresponds to the time interval between the CS and the onset of audience observation. UCS intensity corresponds to the strength of the drive aroused by audience observation, and several specific methods for varying drive strength have been described previously. The duration of the UCS corresponds to the duration of audience observation. Analogs of CS intensity and CS change (stimulus generalization) will be immediately obvious from the ideas already developed.

The strength of the symbolic audience drive may be assessed by its effects in any of the previously discussed drive-testing paradigms, including escape conditioning, avoidance conditioning, punishment, and perhaps most expeditiously by the irrelevant drive method so effectively used in existing social facilitation research.

Predictions

Rules of correspondence have been assumed between the variables of the classical conditioning of learned drives and the variables of social facilitation. Theoretically, the functional relations between the conditioning variables should also hold between the corresponding social facilitation variables. Thus, the social facilitation of a response by a symbolic audience should be a function of the various classical conditioning operations used in establishing that symbolic audience; for example, (*a*) an increasing function of the number of classical trials on which the CS has acquired symbolic audience capacity by preceding audience observation (e.g., Goldstein, 1960; Kalish, 1954); (*b*) an increasing function of the strength of the drive aroused by audience observation (Goldstein, 1960); (*c*) a multiplicative function of the interaction of the number of classical trials on which the CS has acquired symbolic audience capacity by preceding audience observation with the strength of the drive aroused by that observation (e.g., Goldstein, 1960); and (*d*) a decreasing function of the number of trials on which symbolic audience capacity has been extinguished through presentation of the CS unfollowed by audience observation (e.g., Armus, 1960; Kalish, 1954). Moreover, (*e*) symbolic audience capacity should generalize to similar stimuli (e.g., Desiderato, 1964; McAllister & McAllister, 1965).

---

[5] To be sure, Mead rejected classical conditioning as a basis of symbol formation (e.g., pp. 13, 122), but he had in mind Watsonian crudities and not the Hullian pure stimulus acts which his disciple and editor, Morris (1946), found so acceptable. Nonetheless, a more purely Meadian approach to the problems discussed here would surely involve role playing of the audience, rather than classical conditioning to response-produced cues.

Even under ideal conditions in the rat laboratory, higher order conditioning effects are frequently delicate and transitory. Researchers may thus require some of the patience and thoroughness characteristic of workers in the conditioning laboratory. To be sure, higher order conditioning appears to be more easily demonstrated with learned drives than with any other response system (e.g., Anderson et al., 1969; Kamil, 1969), but it is possible that this effect is more of a tribute to the power of electric shock as a primary UCS than it is to the conditionability of drives per se.

Classical Conditioning of Attitudes

Using the evaluative meaning scale of the semantic differential as a measure of attitude, Staats and Staats (e.g., 1958; Staats, Staats, & Heard, 1960) have shown that it is possible to classically condition positive and negative attitudes toward verbal or other stimuli. This procedure is, of course, not one of the standard drive-testing paradigms since it cannot be applied to nonarticulate organisms. It seems certain, however, that this technique will come to be seen as one of the most important tools of researchers in human learning and human social learning, since it is ideally suited for approaching a host of problems, including those addressed in this paper. Using previously neutral words as conditioned stimuli in the symbolic audience procedure, and negative evaluative meaning responses on the semantic differential as a measure of the classical CR, it should be possible to directly apply the above rules of correspondence and the classical conditioning model to the conditioning of attitudes toward audience observation. This attitude conditioning technique would permit us to measure symbolic audience phenomena in an early stage, since the words which have acquired emotional meaning may then be used as punishers or motivators (Staats, 1968).

Conclusion

Present research on the drive theory of social facilitation has been unnecessarily limited by restricting itself to only one of the several drive-testing paradigms used in learning research. These studies are very valuable, but the full possibilities inherent in the drive theory of social facilitation will not be realized until other drive-testing paradigms are also studied. Extension of the drive theory of social facilitation to include such problems as escape and avoidance of audience observation, as well as the acquisition of symbolic audiences, makes it possible to recognize and study more of the great richness of phenomena inherent in one of the traditional central problems of social psychology.

REFERENCES

Amoroso, D. M., & Walters, R. H. Effects of anxiety and socially mediated anxiety reduction on paired-associated learning. *Journal of Personality and Social Psychology*, 1969, **11**, 388–396.

Amsel, A. Partial reinforcement effects on vigor and persistence. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation.* Vol. 1. New York: Academic Press, 1967.

Anderson, D. C., Johnson, D., & Kempton, H. Second-order fear conditioning as revealed through augmentation of a startle response: Part II. *Psychonomic Science*, 1969, **16**, 7–9.

Anderson, N. H. Comparison of different populations: Resistance to extinction and transfer. *Psychological Review*, 1963, **70**, 162–179.

Appel, J. B., & Peterson, N. J. Punishment: Effects of shock intensity on response suppression. *Psychological Reports*, 1965, **16**, 721–730.

Armus, H. L. Effect of percentage of reinforcement and distribution of acquisition trials on extinction of conditioned fear. *Psychological Reports*, 1960, **6**, 387–390.

Armus, H. L., Carlson, K. R., Guinan, J. F., & Crowell, R. A. Effect of a secondary reinforcement stimulus on the auditory startle response. *Psychological Reports*, 1964, **14**, 535–540.

Armus, H. L., & Sniadowski-Dolinsky, D. Startle decrement and secondary reinforcement stimulation. *Psychonomic Science*, 1966, **4**, 175–176.

Aronfreed, J. *Conduct and conscience.* New York: Academic Press, 1968.

Berger, S. M., & Lambert, W. W. Stimulus-response theory in contemporary social psychology. In G. Lindzey & E. Aronson (Eds.), *The handbook of social psychology.* Vol. 1. (2nd ed.) Reading, Mass.: Addison-Wesley, 1968.

Bovard, E. The effects of social stimuli on the response to stress. *Psychological Review*, 1959, **66**, 267–277.

Bower, G. H. Partial and correlated reward in escape learning. *Journal of Experimental Psychology*, 1960, **59**, 126–130.

Bower, G. H., Fowler, H., & Trapold, M. A. Escape learning as a function of amount of shock reduction. *Journal of Experimental Psychology*, 1959, **58**, 482–484.

Brodbeck, M. Models, meaning and theories. In L. Gross (Ed.), *Symposium on sociological theory.* New York: Harper & Row, 1959.

Brown, J. S., & Farber, I. E. Secondary motivational system. *Annual Review of Psychology*, 1968, **19**, 99–134.

Bull, J. A., III, & Overmier, J. B. Additive and subtractive properties of excitation and inhibition. *Journal of Comparative and Physiological Psychology*, 1968, **66**, 511–514.

Bull, J. A., III, & Overmier, J. B. Incompatibility of appetitive and aversive conditioned motivation. *Proceedings of the 74th Annual Convention of the American Psychological Association*, 1969, **4**, 97–98.

Byrne, D. Attitudes and attraction. In L. Berkowitz (Ed.), *Advances in experimental social psychology.* Vol. 4. New York: Academic Press, 1969.

Campbell, B. A., & Kraeling, D. Response strength as a function of drive level and amount of drive reduction. *Journal of Experimental Psychology*, 1953, **45**, 97–101.

Campbell, N. R. *Physics, the elements.* Cambridge, England: Cambridge University Press, 1920.

Cottrell, N. B. Performance in the presence of other human beings: Mere presence, audience and affiliation effects. In E. C. Simmel, R. A. Hoppe, & G. A. Milton (Eds.), *Social facilitation and imitative behavior.* Boston: Allyn & Bacon, 1968.

Cottrell, N. B., Rittle, R. H., & Wack, D. L. The presence of an audience and list type (competitional or noncompetitional) as joint determinants of performance in paired associate learning. *Journal of Personality*, 1967, **35**, 425–434.

Cottrell, N. B., Wack, D. L., Sekerak, G. J., & Rittle, R. H. Social facilitation of dominant responses by the presence of an audience and the mere presence of others. *Journal of Personality and Social Psychology*, 1968, **9**, 245–250.

Cravens, R. W., & Renner, K. E. Conditioned hunger. *Journal of Experimental Psychology*, 1969, **81**, 312–316.

Daly, H. B. Learning of a hurdle-jump response to escape cues paired with reduced reward or frustrative nonreward. *Journal of Experimental Psychology*, 1969, **79**, 146–157.

D'Amato, M. R. Instrumental conditioning. In M. H. Marx (Ed.), *Learning: Processes.* London: Macmillan, 1969.

Davitz, J. R., & Mason, D. J. Socially facilitated reduction of a fear response in rats. *Journal of Comparative and Physiological Psychology*, 1955, **48**, 149–151.

Desiderato, O. Generalization of acquired fear as a function of CS intensity and number of acquisition trials. *Journal of Experimental Psychology*, 1964, **67**, 41–47.

Dinsmoor, J. A. Escape from shock as a conditioning technique. In M. R. Jones (Ed.), *Miami symposium on the prediction of behavior, 1967: Aversive stimulation.* Coral Gables: Miami University Press, 1968.

Dollard, J., & Miller, N. E. *Personality and psychotherapy.* New York: McGraw-Hill, 1950.

Fowler, H., & Trapold, M. A. Escape performance as a function of delay of reinforcement. *Journal of Experimental Psychology*, 1962, **63**, 464–467.

Franchina, J. J. Effects of shock schedules on the acquisition and extinction of escape behavior. *Psychonomic Science*, 1966, **4**, 277–278.

Franchina, J. J. Intertrial intervals and shock schedules in escape training. *Journal of Comparative and Physiological Psychology*, 1969, **67**, 510–515. (a)

Franchina, J. J. Escape behavior and shock intensity: Within-subject versus between-groups comparisons. *Journal of Comparative and Physiological Psychology*, 1969, **69**, 241–245. (b)

Goldstein, M. L. Acquired drive strength as a joint function of shock intensity and number of acquisition trials. *Journal of Experimental Psychology*, 1960, **60**, 349–358.

Henchy, T., & Glass, D. C. Evaluation apprehension and the social facilitation of dominant and subordinate responses. *Journal of Per-

sonality and Social Psychology, 1968, **10**, 446–454.

Hull, C. L. *Principles of behavior.* New York: Appleton-Century-Crofts, 1943.

Hull, C. L. *A behavior system.* New Haven: Yale University Press, 1952.

Jones, M. R. An experimental study of extinction. *Psychological Monographs,* 1953, **67**(19, Whole No. 369).

Kalish, H. I. Strengths of fear as a function of the number of acquisition and extinction trials. *Journal of Experimental Psychology,* 1954, **47**, 1–9.

Kamil, A. C. Some parameters of the second-order conditioning of fear in rats. *Journal of Comparative and Physiological Psychology,* 1969, **67**, 364–369.

Kamin, L. J. The delay-of-punishment gradient. *Journal of Comparative and Physiological Psychology,* 1959, **52**, 434–437.

Karsh, E. B. The effects of number of rewarded trials and intensity of punishment on running speed. *Journal of Comparative and Physiological Psychology,* 1962, **55**, 44–51.

Kiesler, S. B. Stress, affiliation and performance. *Journal of Experimental Research in Personality,* 1966, **1**, 227–235.

Klinger, E. Feedback effects and social facilitation of vigilance performance: Mere coaction versus potential evaluation. *Psychonomic Science,* 1969, **14**, 161–162.

Lachman, R. The model in theory construction. *Psychological Review,* 1960, **67**, 113–129.

Logan, F. A. The Hull-Spence approach. In S. Koch (Ed.), *Psychology: A study of a science.* Vol. 2. *General systematic formulations, learning, and special processes.* New York: McGraw-Hill, 1959.

Logan, F. A. *Incentive.* New Haven: Yale University Press, 1960.

Lott, A. J., & Lott, B. E. A learning theory approach to interpersonal attitudes. In A. G. Greenwald, T. C. Brock, & T. M. Ostrom (Eds.), *Psychological foundations of attitudes.* New York: Academic Press, 1968.

Martens, R. Effects of an audience on learning and performance of a complex motor skill. *Journal of Personality and Social Psychology,* 1969, **12**, 252–260. (a)

Martens, R. Palmar sweating and the presence of an audience. *Journal of Experimental Social Psychology,* 1969, **5**, 371–374. (b)

Matlin, M. W., & Zajonc, R. B. Social facilitation of word associations. *Journal of Personality and Social Psychology,* 1968, **10**, 455–460.

McAllister, D. E., & McAllister, W. R. Second-order conditioning of fear. *Psychonomic Science,* 1964, **1**, 383–384.

McAllister, W. R., & McAllister, D. E. Variables influencing the conditioning and measurement of acquired fear. In W. F. Prokasy (Ed), *Classical conditioning.* New York: Appleton-Century-Crofts, 1965.

McAllister, W. R., & McAllister, D. E. Drive and reward in aversive learning. *American Journal of Psychology,* 1967, **80**, 377–383.

Mead, G. H. *Mind, self and society.* (Posthumously edited by C. W. Morris) Chicago: University of Chicago Press, 1934.

Melvin, K. B., & Brown, J. S. Neutralization of an aversive light stimulus as a function of the number of paired presentations with food. *Journal of Comparative and Physiological Psychology,* 1964, **58**, 350–353.

Miller, N. E. Studies of fear as an acquirable drive: I. Fear as motivation and fear-reduction as reinforcement in the learning of new responses. *Journal of Experimental Psychology,* 1948, **38**, 89–101.

Miller, N. E. Liberalization of basic S-R concepts: Extensions to conflict behavior, motivation and social learning. In S. Koch (Ed.), *Psychology: A study of a science.* Vol. 2. *General systematic formulations, learning, and special processes.* New York: McGraw-Hill, 1959.

Miller, N. E. Learning resistance to pain and fear: Effects of overlearning, exposure, and rewarded exposure in context. *Journal of Experimental Psychology,* 1960, **60**, 137–145.

Miller, N. E., & Dollard, J. *Social learning and imitation.* New Haven: Yale University Press, 1941.

Morris, C. *Signs, language and behavior.* New York: Prentice-Hall, 1946.

Morrison, B. J., & Thatcher, K. Overpopulation effects on social reduction of emotionality in the albino rat. *Journal of Comparative and Physiological Psychology,* 1969, **69**, 658–662.

Mowrer, O. H., & Viek, P. An experimental analogue of fear from a sense of helplessness. *Journal of Abnormal and Social Psychology,* 1948, **43**, 193–200.

Myers, A. K., & Miller, N. E. Failure to find a learned drive based on hunger: Evidence for learning motivated by "exploration." *Journal of Comparative and Physiological Psychology,* 1954, **47**, 428–436.

Nelson, F. Effects of two counterconditioning procedures on the extinction of fear. *Journal of Comparative and Physiological Psychology,* 1966, **62**, 208–213.

Notterman, J. M., & Mintz, D. E. *Dynamics of response.* New York: Wiley, 1965.

Novin, D., & Miller, N. E. Failure to condition thirst induced by feeding dry food to hungry rats. *Journal of Comparative and Physiological Psychology,* 1962, **55**, 373–374.

Oppenheimer, R. Analogy in science. *American Psychologist,* 1956, **11**, 127–135.

Paul, G. L. Insight versus desensitization in psychotherapy two years after termination. *Journal of Consulting Psychology,* 1967, **31**, 333–348.

Paulus, P. B., & Murdoch, P. Anticipated evaluation and audience presence in the enhancement of dominant responses. *Journal of Experimental Social Psychology*, in press.

RAZRAN, G. Conditioning away social bias by the luncheon technique. *Psychological Bulletin*, 1938, **35**, 693.

RESCORLA, R. A. Conditioned inhibition of fear resulting from negative CS-US contingencies. *Journal of Comparative and Physiological Psychology*, 1969, **67**, 504–509.

RESCORLA, R. A., & LOLORDO, V. M. Inhibition of avoidance behavior. *Journal of Comparative and Physiological Psychology*, 1965, **59**, 406–412.

SKINNER, B. F. Operant behavior. *American Psychologist*, 1963, **18**, 503–515.

SPENCE, J. T., & SPENCE, K. W. The motivational components of manifest anxiety: Drive and drive stimuli. In C. D. Spielberger (Ed.), *Anxiety and behavior*. New York: Academic Press, 1966.

SPENCE, K. W. Behavior theory and conditioning. New Haven: Yale University Press, 1956.

SPENCE, K. W. Cognitive and drive factors in the extinction of the conditioned eye blink in human subjects. *Psychological Review*, 1966, **73**, 445–458.

SPENCE, K. W., HAGGARD, B. F., & ROSS, L. E. UCS intensity and associative (habit) strength in the eyelid CR. *Journal of Experimental Psychology*, 1958, **55**, 404–411.

STAATS, A. W. Social behaviorism and human motivation: Principles of the attitude-reinforcer-discriminative system. In A. G. Greenwald, T. C. Brock, & T. M. Ostrom (Eds.), *Psychological foundations of attitudes*. New York: Academic Press, 1968.

STAATS, A. W., & STAATS, C. K. Attitudes established by classical conditioning. *Journal of Abnormal and Social Psychology*, 1958, **57**, 37–40.

STAATS, C. K., STAATS, A. W., & HEARD, W. G. Attitude development and ratio of reinforcement. *Sociometry*, 1960, **23**, 338–350.

THIBAUT, J. W., & KELLEY, H. H. *The social psychology of groups*. New York: Wiley, 1959.

TRAPOLD, M. A. The effect of incentive motivation on an unrelated reflex response. *Journal of Comparative and Physiological Psychology*, 1962, **55**, 1034–1039.

TRAPOLD, M. A., & FOWLER, H. Instrumental escape performance as a function of the intensity of noxious stimulation. *Journal of Experimental Psychology*, 1960, **60**, 323–326.

TRAPOLD, M. A., & SPENCE, K. W. Performance changes in eyelid conditioning as related to the motivational and reinforcing properties of the UCS. *Journal of Experimental Psychology*, 1960, **59**, 209–213.

TRAPOLD, M. A., & WINOKUR, S. Transfer from classical conditioning and extinction to acquisition, extinction, and stimulus generalization of a positively reinforced instrumental response. *Journal of Experimental Psychology*, 1967, **73**, 517–525.

WAGNER, A. R. Conditioned frustration as a learned drive. *Journal of Experimental Psychology*, 1963, **66**, 142–148.

WEISS, J. M. Effects of coping responses on stress. *Journal of Comparative and Physiological Psychology*, 1968, **65**, 251–260.

WEISS, R. F. Response speed, amplitude, and resistance to extinction as joint functions of work and length of behavior chain. *Journal of Experimental Psychology*, 1961, **61**, 245–256.

WEISS, R. F. Persuasion and the acquisition of attitudes: Models from conditioning and selective learning. *Psychological Reports*, 1962, **11**, 709–732.

WEISS, R. F. Defection from social movements and subsequent recruitment to new movements. *Sociometry*, 1963, **26**, 1–20.

WEISS, R. F. An extension of Hullian learning theory to persuasive communication. In A. G. Greenwald, T. C. Brock, & T. M. Ostrom (Eds.), *Psychological foundations of attitudes*. New York: Academic Press, 1968.

WEISS, R. F., MILLER, F. G., LANGAN, C. J., & CECIL, J. S. Social facilitation of attitude change. *Psychonomic Science*, 1971, in press.

WHEELER, L. Toward a theory of behavioral contagion. *Psychological Review*, 1966, **73**, 179–192.

WHITING, J. W. M., & CHILD, I. L. *Child training and personality: A cross-cultural study*. New Haven: Yale University Press, 1953.

WILSON, E. H., & DINSMOOR, J. A. Effect of feeding on "fear" as measured by passive avoidance in rats. *Journal of Comparative and Physiological Psychology*, 1970, **70**, 431–436.

WOLPE, J. *Psychotherapy by reciprocal inhibition*. Stanford: Stanford University Press, 1958.

WOODS, P. J., DAVIDSON, E. H., & PETERS, R. J. Instrumental escape conditioning in a water tank: Effects of variations in drive stimulus intensity and reinforcement magnitude. *Journal of Comparative and Physiological Psychology*, 1964, **57**, 466–470.

WOODS, P. J., & FELDMAN, G. B. Combination of magnitude and delay of reinforcement in instrumental escape conditioning. *Journal of Comparative and Physiological Psychology*, 1966, **62**, 149–151.

ZAJONC, R. B. Social facilitation. *Science*, 1965, **149**, 269–274.

ZAJONC, R. B., HEINGARTNER, A., & HERMAN, E. M. Social enhancement and impairment of performance in the cockroach. *Journal of Personality and Social Psychology*, 1969, **13**, 83–92.

(Received April 8, 1970)

Psychological Review
1971, Vol. 78, No. 1, 58–70

PUNISHMENT:

METHOD AND THEORY[1]

PHILIP J. DUNHAM[2]

Dalhousie University

A methodological framework for the analysis of punishment is outlined. The methodology, which is called a multiple-response base-line procedure, serves two purposes. First, it raises a number of new questions about the properties of punishment. Second, it permits the examination of some untested assumptions found in traditional punishment theory. Initial evidence obtained with the multiple-response methodology questions the validity of traditional theoretical assumptions and suggests two simple rules for predicting the properties of various punishment operations.

When an aversive stimulus is contingent upon the occurrence of a particular response, a decrement in the probability of the response is usually observed. This procedure is typically called punishment and the decrement in response probability is called punishment suppression. The basic purpose of this paper is to delineate some fundamental problems with existing punishment theory and to suggest an alternative approach to the problem of punishment.

An Overview of Punishment Theory

Historically, there have been two fundamental assumptions used to explain the phenomenon of punishment suppression. The first of these assumptions was the strong version of the negative Law of Effect proposed by Thorndike (1913). Thorndike assumed that any painful or unpleasant event would weaken the response (or assumed S–R bond) which preceded that event. Thorndike (1932) subsequently rejected this notion and it has not enjoyed any serious attention since that time. The second fundamental assumption suggested to account for the punishment suppression phenomenon has been referred to as the *alternative-response assumption* (cf. Dunham, Mariner, & Adams, 1969). In its simplest form, the assumption states that the decrement in a punished response is caused by an increment in some alternative behavior.

All contemporary explanations of punishment suppression are specific elaborations of this alternative-response assumption. Those specific elaborations which have been most formalized fall into two major categories. These categories are referred to as single-process and two-process theories of punishment (cf. Solomon, 1964). The earmark of the single-process theory is the assumption that only one type of learning mechanism is involved in the development and maintenance of the alternative response during punishment training. Two types of single-process theory have been suggested and are differentiated in terms of suggesting either a classical or an instrumental conditioning mechanism. Estes and Skinner (1941), for example, suggested that emotional responses elicited by the punishing event are *classically* conditioned to stimuli which precede the punishing event. The classically conditioned behavior is assumed to compete with the punished response and cause the suppression. Miller and Dollard (1941) exemplify the instrumental conditioning version of single-process theory. They suggested that any response which is associated with the termination of the punishing stimulus will be *instrumentally* conditioned as a response which escapes pain and competes directly with the punished response.

---

[1] Research reported in this paper was supported by Project Grant APA-194 from the National Research Council of Canada. Many valuable suggestions have been made by different individuals during the preparation of this manuscript. The author is particularly grateful to N. J. Mackintosh, C. J. Brimer, and B. Moore for their critical comments.

[2] Requests for reprints should be sent to Philip J. Dunham, Department of Psychology, Dalhousie University, 1460 Oxford Street, Halifax, Nova Scotia, Canada.

Two-process punishment theories specify two different learning mechanisms which are sequentially involved in the development and maintenance of the assumed alternative response. Dinsmoor (1954, 1955) and Mowrer (1947) are formal examples of the two-process explanation of punishment suppression. Dinsmoor suggested that the proprioceptive stimulus feedback from the punished response acquires secondary aversive properties via *classical* pairings with the primary aversive event. Any response which is *instrumental* in disrupting this chain of conditioned aversive stimulation will develop and be maintained as a response which competes with the punished response. Hence, two learning mechanisms, first classical then instrumental, are assumed to operate in the development of the alternative behavior. Mowrer's (1947) version of two-process theory substitutes the notion of conditioned fear for the notion of aversive stimulation conditioned in the Dinsmoor theory.

There are two basic implications of any version of the alternative response theory. First, there is the implication that some alternative behavior will develop and be maintained during punishment training. Second, there is the implication that this alternative response causes the reduction in the punished response. Presumably the former implication could be confirmed independent of the latter. But the latter could not be confirmed if the former were false.

In spite of the substantial amount of research on punishment in the last decade, there is no *direct* evidence to support either of the above implications of the alternative-response assumption. As Azrin and Holz (1966) have stated, the typical procedure in punishment research has been to infer the presence of the alternative response from the absence of the punished response. A minimal requirement for testing the assumption is the measurement of the alternative response independent of the phenomenon which it seeks to explain.

Two reasons can be suggested for the lack of direct evidence bearing on the alternative-response assumption. In the case of single-process theories, contingencies are specified which make the response elicited by the punishing event a prime candidate for conditioning during the suppression of the punished response. It is not a profound observation to note that punishment procedures which have traditionally been employed involve organisms, aversive stimuli, and apparatus which make it difficult to measure those responses which are suspected to participate in the relevant contingency. These "emotional" behaviors have not been recorded on impulse counters and this has prevented the accumulation of any evidence concerning changes in their probability during punishment training. The major problem with single-process versions of the alternative-response assumption at this point would appear to be the lack of an adequate methodology to test what would appear to be very testable implications.

With respect to the two-process theories, the problem is more serious. As Schuster and Rachlin (1968) have suggested:

> Because both the reinforcer and the response are unobserved and unobservable, the two factor theory of punishment poses a serious problem for the experimenter who wishes to test it: how can it be disproved? All the critical events are assumed to occur within the organism being punished [p. 784].

In addition to the specification of the critical contingencies "inside" the organism, it should be noted that *any* increase in some alternative behavior observed during punishment can be taken as support for the operation of the assumed internal contingency. Hence, the measurement of the unobserved behavior referred to by Schuster and Rachlin does not make the positions any more susceptible to disproof.

The picture which emerges from this brief overview of punishment theory and research is that there has been a lack of interaction between punishment theory and punishment data. The lack of any data relevant to the most fundamental theoretical assumptions has permitted the alternative response interpretations to persist as originally formulated.

In the next section of the discussion, a methodological approach to punishment will be described which can serve two functions in the context of the existing punishment literature. First, the methodology permits us to examine some questions about the properties of punishment which have not previously been considered. Second, it permits us to examine some previously untested implications of existing punishment theory.

A Methodological Approach to Punishment

The methodological approach to be suggested is called a multiple-response base-line procedure. It can be applied to a variety of problems in addition to punishment, and when viewed in the context of traditional methodology, it falls between a typical operant procedure, in which a single response is shaped under constraints, and typical ethological procedures, in which behavior is observed without external laboratory constraints. The most convenient way to describe the essential features of the methodology is to elaborate on a specific hypothetical example which is representative of several obvious variations on the basic approach. The hypothetical example should be noted with some care since it will be approximated in reality when experimental evidence is subsequently discussed.

Consider a small animal chamber with grid floor and three sources of enjoyment for the small rodent commonly called a Mongolian gerbil (*Meriones unquiculates*). The three items of interest in the chamber are a food bin with an unlimited supply of standard Noyes pellets, a drinking tube with unlimited supply of water, and adding machine paper which is threaded through a slot in the wall of the chamber.

The reader familiar with the behavior of this curious rodent will not be surprised to find that the gerbil will spend much of a half-hour daily session in the chamber shredding the adding machine paper, eating the food pellets, and drinking from the tube, in that order of preference. It is relatively easy to record the duration of each of these three behaviors during the session and convert these duration measures to a scale of response probability by dividing the amount of time spent in a particular response state by the total time possible (cf. Premack, 1965). Any type of behavior can be measured in terms of its duration, including the class of behavior which is labeled "doing nothing." With the appropriate manipulanda for a particular organism and control of the relevant parameters, a multiple-response steady-state base line of behavior, in which several measurable responses are observed to fill experimental time, can be established. No shaping or active manipulation of contingencies is assumed to be necessary. Assume, for the purposes of discussion, that our gerbil cooperated and filled the half hour with drinking ($p = .1$), eating ($p = .3$), and paper shredding ($p = .6$).

Once the multiple-response base line has stabilized, we are in the interesting position of actively manipulating the organism's environment and assessing the effects of such manipulations on *all* of the responses in the repertoire. Obviously, the manipulations of most interest in the present context are those which we call punishment operations; however, it is instructive to digress briefly and consider a simple manipulation like making a running wheel available. The most visible effect of making the running wheel available will be introduction of a running response into a repertoire of behavior which already fills experimental time. If running occurs, by definition, there will be a decrement in the observed probability of one or more of the existing responses in the repertoire. Curiously, there is little more than intuition to tell us how the organism will reorganize his response hierachy to accommodate the running response. Will he sacrifice a little bit of each of the existing behaviors? Will he select one response and sacrifice it for running privileges? If the latter, what is the rule of response selection?

If an aversive stimulus like electric shock is introduced, one is faced with roughly the same problem as that posed by the introduction of the running wheel. Shock, as an unconditioned stimulus, will define a certain probability of unconditioned behavior which must be assimilated into a response hierarchy. By definition, some decrement in the probability of one or more of the existing responses must take place. Again, we are not sure how the organism alters the existing preference structure to accomodate this additional behavior, but the rules which describe such alterations would be of importance in predicting the *immediate* effects of any variety of punishment operation on responses in the repertoire.

The point to be made with the two preceding examples is that some manipulations which can be used in the context of a multiple-response base line have the property of adding a response to the existing repertoire of behavior. Punishment procedures are one such manipulation, and it is suggested that procedures be arranged which permit the a priori specification of the response which will be added and subsequent measurement of that response during punishment training. In this respect, the selection of the gerbil was fortuitous. The response which we have observed to be associated with the introduction of shock is a vigorous biting and chewing of the grid floor. As reliable, if not as desirable, is an aggressive attack on another gerbil if the target animal is provided.

When one considers the variety of ways in which an aversive event like electric shock can be introduced into a multiple-response procedure, a number of empirical questions are generated which have not received previous experimental attention. A brief consideration of a few of these questions will illustrate the heuristic value of the multiple-response base-line procedure.

The traditional literature dealing with shock provides us with two major classes of operation which are called punishment and avoidance, with variations on each theme. In the context of the multiple-response base-line procedure, an alternative operational dichotomy is suggested. Specifically, the shock event can be delivered according to a program which makes reference to one or more responses in the repertoire, or according to a program without reference to the organism's response repertoire. The former includes traditional response-contingent punishment and Sidman avoidance operations; the latter includes various temporal schedules which are arranged independent of the organism's behavior. In future discussion, the terms response contingent and noncontingent will be used to describe these generic operational categories. Some measure of the comprehensiveness of any theoretical framework is provided by its ability to subsume data in both operational categories. Several of these operations, most often employed in single-response operant research, reveal interesting new dimensions when considered in the multiple-response context.

Consider the response-contingent punishment operation where one response is selected from the repertoire and the onset of that response is followed immediately by an instance of shock. The suppressive effects of this operation on the referent behavior are well known. However, the operation defines a version of Sidman avoidance contingency for all other responses in the organism's repertoire, including the response elicited by the shock event. In the gerbil example, response-contingent shock for eating also defines a very effective Sidman avoidance contingency for paper shredding. The more time spent paper shredding, the longer the interval between shocks and the fewer shocks received. Does the gerbil adjust his behavior in a manner suggested by the contingency?

Consider a fixed-interval noncontingent shock procedure. For a given density of shock, the probability of each response in the organism's repertoire will determine the number of shocks which are associated with that particular response. In the gerbil example, paper shredding occupies over half the experimental session, hence more shocks will arrive during paper shredding than during any other response in the repertoire. Of more interest, with shocks delivered at fixed intervals of time, that response which most consistently follows the shock event will most consistently predict the longest interval of "safety." Again, one must ask if the organism changes his behavior according to these implicit contingencies defined by the fixed-interval noncontingent schedule. A phenomenon reported initially by Morse, Mead, and Kelleher (1967) is perhaps relevant. These investigators reported that monkeys which are exposed to fixed-interval noncontingent shock will aggressively bite a rubber tube made available after each shock is delivered. The postshock elicited biting behavior would be the response in the organism's repertoire which most consistently predicts the longest interval of safety. As suggested by the preceding analysis, the animal changes its biting behavior accordingly. The amount of biting behavior increases during extended punishment training and eventually fills the shock–shock interval with a "scallop" in the rate of biting which appears at the end of the interval.

Consider also a Sidman avoidance procedure with a constant shock–shock interval and constant response–shock interval. Once we select a single response from the multiple-response base line as an avoidance response, all other responses in the repertoire are initially punished according to a noncontingent shock schedule, with the shortest possible intershock interval being the Sidman shock–shock interval. Which of the responses in the organism's repertoire should one select for the avoidance response for efficient learning? It would seem obvious that the response elicited by the shock would enjoy some advantages not enjoyed by the other responses in the repertoire. First, that response will predict the longest safe interval between shocks even if the avoidance contingency were not in effect. Second, it has a relatively high probability during initial training sessions, hence it will sample an implicit avoidance contingency quite often. Third, if it is explicitly given the avoidance property of delaying shock, there should be few responses which will compete with it for rapid learning. This suggestion is in line with Bolles' (1970) view that the organism's innate species-specific defense reactions (SSDRs) to an aversive stimulus are most readily learned as avoidance responses. I would suggest, however, that the rapid learning has nothing to do with their "innate defensive" properties. It is the property of immediately following the shock event which optimizes the conditions for avoidance learning—that is, not only does the response eventually reduce shock frequency, but it initially predicts safety for a longer period of time than any other response in the organism's repertoire.

The converse situation can also be suggested. If we give the elicited response the initial advantage of consistently predicting the longest safe interval of time, yet attempt to train a different response as an avoidance behavior, it would not be surprising to observe the animal attempting initially to develop the elicited behavior to the detriment of the avoidance response. This would be particularly true in the case of constant shock–shock intervals (e.g., the Morse, Mead, & Kelleher, 1967, phenomenon), as opposed to variable shock–shock intervals where the predictive properties of the elicited response are less evident. This also leads to the suggestion that a variable shock–shock interval will produce faster learning of avoidance behavior when the avoidance response is *other than the shock-elicited response* (cf. Bolles & Popp, 1964).

Finally, it is of some interest to consider the variations in temporal schedule which are possible with noncontingent shock. We have already discussed the fixed-interval case; now consider some implicit contingencies established by variable-interval cases. When programming shocks to occur at varying intervals in time, the distribution of intervals may be an important consideration. The typical variable-interval tape program is a rectangular distribution of intervals with some guaranteed minimum interval. It is typically described in terms of its arithmetic mean and delivers the shocks with a random sequence of intershock intervals. Of more interest is an exponential distribution of intervals which will deliver the same number of shocks at varying intervals, but is a continuous distribution. The probability of shock at any "moment" in time is a fixed value and the animal has no programmed interval which is guaranteed to be safe. In typical punishment studies where a single operant response is shaped, trained, and measured, the effects of shocks delivered according to fixed intervals, rectangular distributions of variable intervals, and exponential distributions of variable intervals be the same—the operant is suppressed. However, in the multiple-response procedure, very different effects may be found with the three distributions. As mentioned earlier, the elicited response enjoys the advantage of predicting a safe period most consistently in the fixed-interval shock schedule. This would be true to a lesser degree of variable-interval schedules with rectangular distributions. The elicited response would most consistently predict the minimum interval of safety programmed in that distribution. In the exponential distribution, however, no response consistently predicts a minimum safe interval. According to the program, shock is equally probable at any point in time. This removes those implicit contingencies for certain responses which are most evident in fixed-interval schedules. Again the question which must be asked is whether or not the organism adjusts his behavior according to the presence and absence of these implicit contingencies.

The preceding examples are intended to illustrate two points: first, the heuristic value of the multiple-response analysis in terms of generating testable questions; second, the sparsity of evidence relevant to these questions. This sparsity is understood when one recognizes the emphasis which has traditionally been placed on single-response measurement in both free operant and discrete-trial punishment research.

The discussion of the multiple-response methodology can be concluded with a brief consideration of the predictions which single-process and two-process punishment theories would make in the context of the hypothetical gerbil procedure. With respect to single-process theory, the use of any type of punishment operation will introduce the grid-biting response into the organism's response repertoire. This grid-biting response should be a prime candidate for both classical conditioning (Estes & Skinner, 1941) and instrumental escape conditioning (Miller & Dollard, 1941) since it follows the shock event more often than other responses. The predictions of the two-process versions of the alternative-response assumption must be considered separately in terms of predictions about the effects of contingent and noncontingent shock procedures. Basically, the two-process theories suggest that any response which disrupts the chain of discriminative stimulus conditions which precede the shock event will increase in probability and be maintained during punishment training. In the case of response-contingent punishment, these theories make the general prediction that any response other than the punished response might increase in probability. They do not provide us with a response-selection rule to tell us if one or all of the unpunished responses increases. In the case of noncontingent shock operations, the two-process theories fail to make any meaningful prediction. When a noncontingent shock operation is employed, every response in the organism's response repertoire will be part of a discriminative chain of stimuli which terminates with shock on some occasions. Hence, it is impossible to suggest that any response in the repertoire participates in an instrumental contingency which disrupts the chain of conditioned aversive stimulation.

Some Rules for Prediction

The predictions which the single-process and two-process explanations of punishment suppression make in the context of the multiple-response methodology have been discussed in the preceding section. Prior to considering some evidence, I would like to suggest some alternative rules for predicting the effects of a variety of punishment operations on the various behaviors measured in the multiple-response procedure. The rules are, at this point, tentative, and a systematic analysis of punishment in the context of a multiple-response base line may modify or contraindicate these initial suggestions.

Once the immediate change in performance caused by the introduction of an unconditioned behavior into the repertoire ha taken place (shock is introduced), the organism has a hierarchy of responses described in terms of response probability. Th two rules which attempt to predict the fat of each of the responses in the organism repertoire during subsequent aversive trai ing are the following: (*a*) That particul response in the organism's repertoire whic is most frequently associated with shock onset and/or predicts the onset of shock within a shorter time than other responses will decrease in probability and remain below its operant base line. (*b*) That particular response in the organism's repertoire which is most frequently associated with the absence of shock onset and/or predicts the absence of shock onset for a longer period of time than other responses will increase in probability and remain above its operant base line.

Consider the spirit in which these two rules are formulated. I am suggesting that there are two basic contingencies of importance in any operation for delivering the shock. First, there is an instrumental punishment contingency. It has two important dimensions. A response can be more frequently associated with shock onset than other responses, or it can predict a given frequency of shock onset within a shorter period of time, or both. Thus, the two dimensions of the contingency are frequency and time. Second, there is an instrumental avoidance contingency. It also has two important dimensions. A response can be more frequently associated with the absence of shock onset, or can predict the absence of shock onset for a longer period of time, or both.

In traditional single-response operant methodology, the time not spent on the operant is usually assigned the label of "nonresponse" and is assumed to be a homogeneous mass of behaviors, (cf. Rachlin & Herrnstein, 1969). In spite of their tentative nature, the two rules described above should make it obvious that it may be of more value to recognize the "heterogeneity" of the nonreponse class—at least in terms of response probability differences. Further empirical analysis may reveal such factors as the sequential dependencies between different responses to be an important determinant of the organism's adjustment to aversive contingencies. This criticism of the single-response operant approach is equally applicable to the procedures which employ appetitive contingencies.

Some Evidence

Over the past year, the research consists of a series of successive approximations to the multiple-response methodology which has been discussed throughout this paper. Many of the thoughts developed in the preceding discussion of punishment originated with a serendipitous observation made while studying the effects of punishment on key pecking in pigeons (Dunham et al., 1969). Dunham et al. observed that pigeons trained to peck a response key for grain on a variable-interval schedule missed the key and hit the wall area adjacent to the key with a certain number of pecks. The introduction of key-peck response-contingent punishment suppressed key pecking and increased the frequency of off-key pecks during punishment. These results suggested that an unpunished response, other than the shock-elicited response, will increase in probability and be maintained during punishment training.

Subsequently, a more deliberate approximation to the multiple-response methodology was attempted, taking advantage of the phenomenon of schedule-induced polydipsia (cf. Falk, 1966) in order to establish two steady-state responses which occupied a large portion of experimental time.[3] Falk (1966) reported that rats trained to lever press for food pellets on a variable-interval schedule will indulge in an excessive amount of drinking during an experimental session if a drinking tube is made freely available. Using Falk's standard procedure, two rats were trained one hour each day to the point of polydipsic drinking, and a mild .2-milliampere, .5-second shock was introduced which was contingent on the lever press response (Subject 1) or on the drinking response (Subject 2). With the relatively mild shock intensity to minimize the probability of the unconditioned (and unmeasured) responses elicited by shock, it was assumed that drinking would be the most probable unpunished alternative if lever pressing were punished, and that lever pressing would be the most probable unpunished alternative if drinking were punished. According to the rules outlined in earlier discussion, the punished response was predicted to be suppressed, and the most probable unpunished alternative in the response repertoire was predicted to increase in probability following some degree of initial disruption by the addition of shock-elicited behavior to the response repertoire.

The results of the initial polydipsia training phase, the first punishment training phase, a recovery phase, and a second punishment training phase are illustrated in Figure 1. During initial training, the animals developed polydipsic levels of water intake similar to those usually reported by Falk (1966). In the initial punishment phase, Subject 1 was shocked for lever pressing and Subject 2 was shocked for drinking. In both cases, the punished response was observed to be suppressed immediately and the unpunished alternative behavior was observed, following initial disruption, to increase to levels which exceeded the established prepunishment base-line level. The results obtained with Subject 1 were very surprising. In spite of the aberrant (polydipsic) base line of water intake prior to punishment, the response was observed to exceed that base line during punishment sessions. In the third phase of the procedure, the punishment contingencies were removed and recovery was observed for 14 sessions. With the removal of shock, all responses increased to levels even higher than preceding base lines. This overshoot following removal of punishment is typical of the recovery of punished responses (cf. Azrin & Holz, 1966). It is interesting to note that overshoot occurs with both the punished and unpunished alternatives, which were measured in the present procedure. Following the 14 days of recovery, punishment contingencies were again introduced—this time for the response which had *not* been punished during the first punishment phase. Basically the same phenomena observed during the first punishment phase were observed during the second. The punished responses suppressed and the unpunished alternatives were disrupted and then increased in probability. Unfortunately the increase in drinking observed in Subject 2 did not consistently exceed the prepunishment base line. This may be the result of the ceiling problem on

The final line of evidence from the author's laboratory work to date was an attempt to approximate the hypothetical gerbil procedure used as an example throughout this article. There were several reasons to extend experimentation beyond the case of two highly probable responses to three or more measured responses. First, the tentative rules specify that only one of the several unpublished alternative responses will increase in probability during punishment training. In order to test this prediction, more than one alternative must be measured. Second, a test of single-process theories is possible only if the unconditioned response introduced by the shock event is measured. Third, an alternative approach to the data in the polydipsia experiment would be the suggestion that the shock has some general arousal property which increases all unpublished responding. The demonstration of a single-response increase would make such an arousal mechanism less appealing than the type of contingency analysis suggested by the tentative rules.

To answer these questions, Kennedy Muyesu-Kaisha Munavi conducted a burdensome experiment in the author's laboratory which was a very close approximation to the hypothetical gerbil procedure described earlier. Nine gerbils were randomly assigned to one of three groups. Each gerbil was placed in a small response chamber with food bin, drinking tube, and adding machine paper for a daily half-hour session. The total duration of each of the three behaviors was recorded on elapsed-time meters by an observer looking at the animals through a one-way glass window. When a gerbil entered the chamber, it had been permitted access to only 80% of its norma ad lib intake of laboratory chow during the preceding 23½ hours, and the only wate permitted was that available during the ex perimental session. Under these conditions the gerbils managed to fill approximatel 25 of the 30 minutes available each sessio with paper-shredding, eating, and drinking in that order of preference. It should b noted that records were kept of the amoun of food and water consumed during bas line observations. Subsequently, if eatir

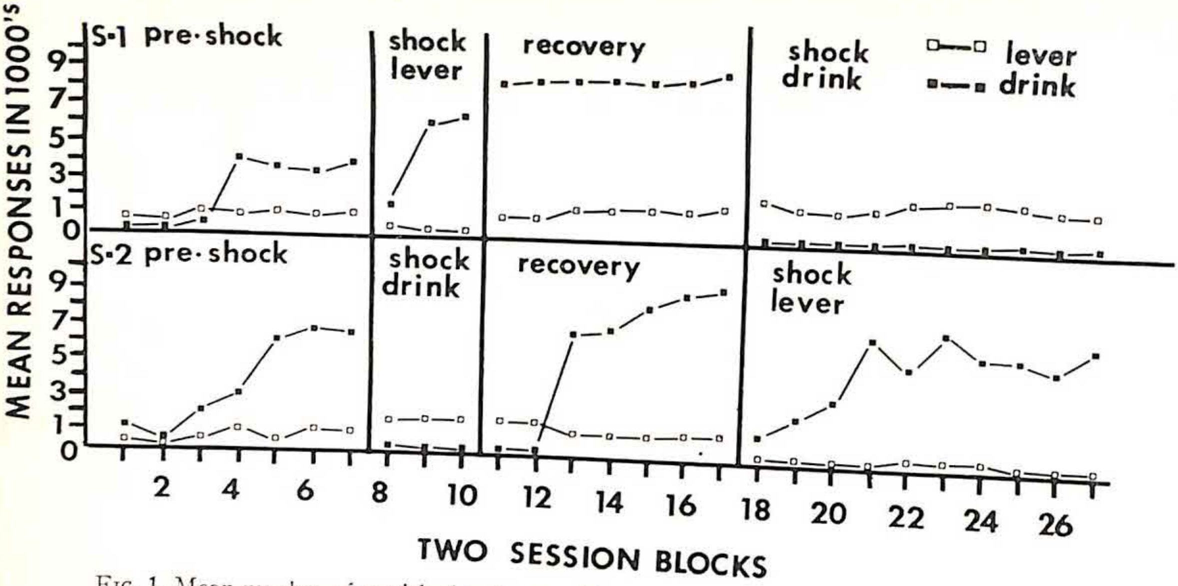


FIG. 1. Mean number of punished and unpunished responses plotted in two-session blocks for Subject 1 (S-1) and Subject 2 (S-2) in the polydipsia procedure.

was the punished behavior, the animal was permitted to make up any deficit in food intake during the experimental session in its daily ration one hour after the experimental session. Hence, there was no confounding of changes in either the total daily food intake or water intake with the introduction of the shock (cf. Dunham & Kilps, 1969).

After a base line of three behaviors was established, Group E was punished for eating, Group P for paper shredding, and Group D for drinking. The shock was a .2-milliampere, .5-second shock delivered at the onset of the referent behavior and continued at 2-second intervals until the referent response was stopped. At this point in the experimentation with gerbils, it was not known what elicited behavior or behaviors would be introduced with this shock intensity. After running four animals through the first session of punishment it was obvious that all animals were attacking the grid bars through which the shock was delivered. At this point a record was started of the duration of grid biting as a response which consistently occurred with the shock in all nine animals. Hence, we have the probability of grid biting during the first punishment session for only five of the nine subjects in this initial experiment. After the first session, grid biting was measured along with the three other responses during each session.

Prior to examining the results of the experiment, it is of some value to consider the predictions made by the rules when applied to this situation: (*a*) the rules suggest that the response most frequently and immediately associated with shock will decrease in probability and remain below its operant base line; it is the referent punished response in the response-contingent procedure; (*b*) there will be an immediate disruption of all responses in the repertoire immediately upon the introduction of the grid-biting response into the three-response repertoire which already nearly fills time; (*c*) after the initial disruption, the response in the repertoire which is most frequently associated with the absence of shock for the longest period of time will increase in probability to levels which exceed the prepunishment base line. This response should be that response which is most probable on the assumption that the most probable behavior will sample the implicit avoidance contingency most often.

Subsidiary predictions are also implicit in the rules. For example, if grid biting is not the most probable unpunished alternative behavior, it will drop out of the repertoire as the shock frequency declines (punished re-

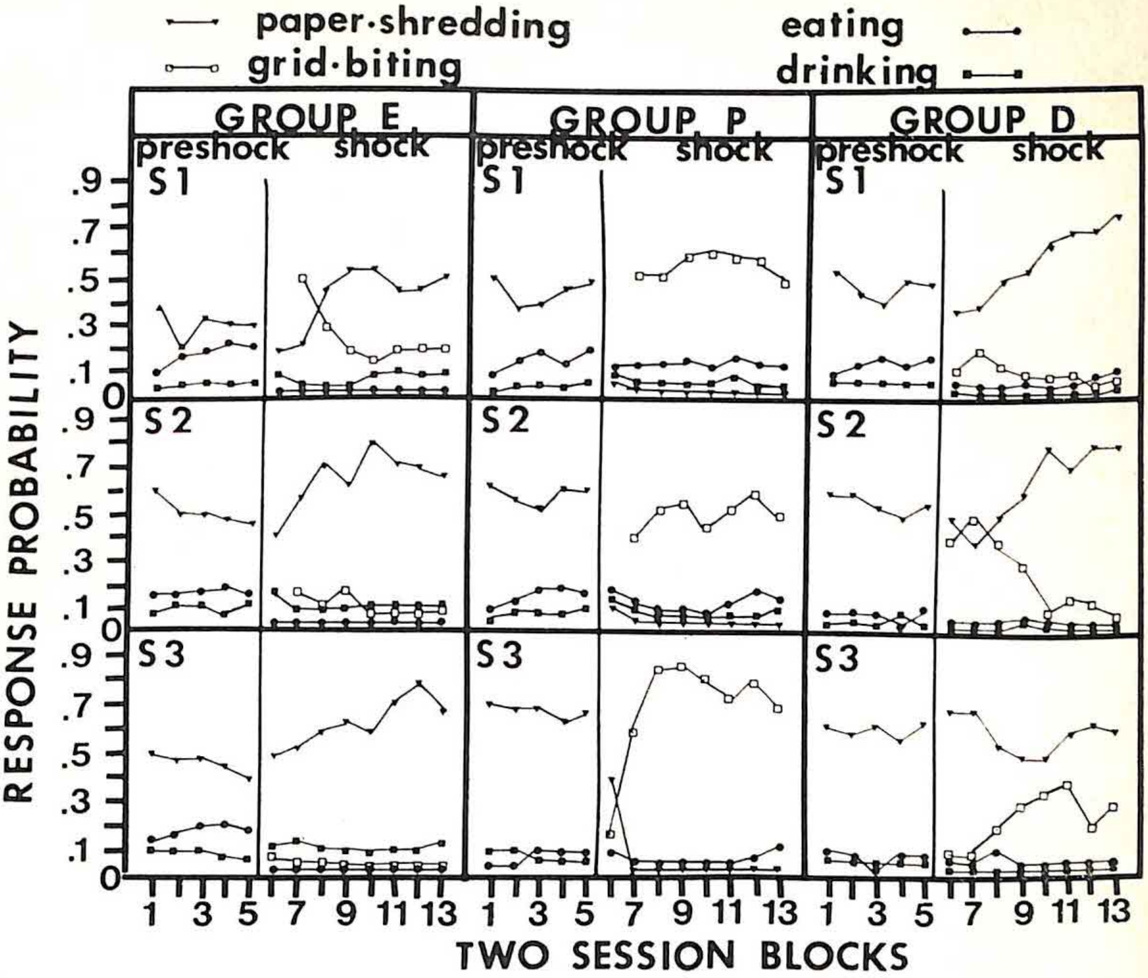


FIG. 2. Probability of punished and unpunished responses for the three subjects in each of the three groups in the gerbil procedure. (Note that all data points are two-session blocks with the exception of the first punishment session which is plotted as a single-session point.)

sponse suppressed). The latter suggestion is contrary to the predictions of the traditional single-process punishment theories. If, however, the grid biting response is the most probable of the unpunished alternatives, it will be maintained even in the absence of the shock UCS during punishment training.

The results of the gerbil experiment for each subject in each of the three groups are presented in Figure 2.

Consider first the two results which are common to all three groups and which are of primary interest in terms of the contingency rules. First, in all cases, the instrumental punishment contingency was sufficient to suppress the referent behavior. Second, and of more interest, in all cases, only *one* of the alternative responses in the repertoire was observed to increase in probability to levels which exceeded the baseline probability. These two results are taken as support for the operation of the contingencies outlined in preceding discussion.

Consider next the results specific to each group. Of particular interest in this context is the response which is selected to increase in probability during the punishment training. The subjects in Group E were punished for eating, and all three of the subjects were observed to increase the probability of paper shredding following an initial period of disruption. The subjects in Group P were punished for paper shredding, and in all three subjects the probability of grid biting increased during punishment training. Group D was punished for drinking and revealed inconsistent results. Subject 1 and Subject 2 increased the probability of paper shredding, while Subject 3 increased the probability of grid biting during punishment.

To what extent are the responses observed to increase in each group predicted by the contingency rules? The rules suggest that the response which most frequently samples the avoidance contingency in early sessions will be the most probable of the safe alternatives. In spite of the fact that the probability of the grid-biting response was not measured during the first session in four of the animals studied, it is safe to assume that two responses were the most probable unpunished behaviors during early punishment sessions. These are grid-biting and paper-shredding responses. Both would sample the avoidance contingency more frequently, and the grid-biting response which followed the shock event would predict the absence of shock for a longer period of time more consistently than the other responses which were not sequentially dependent on the shock event. For these reasons, we would expect all subjects to reveal an increase in either grid biting or paper shredding. However, there were some instances in which grid biting was initially more probable than paper shredding, yet paper shredding was observed to increase to levels above base line as grid biting dropped out (e.g., Group E, Subject 1). Further research, in which more precise measurements are obtained of the degree to which various responses sample both the frequency and time dimensions of the two contingencies, will provide a more critical test of the suggested rules and perhaps suggest refinements.

In addition to suggesting that the methodology and the contingency analysis are tractable approaches to the punishment problem, the gerbil data question the validity of the single-process versions of the alternative-response assumption. Stated quite simply, those cases in the present situation which revealed a gradual decline in the probability of grid biting over the course of punishment training are exactly the opposite to changes expected from traditional single-process assumptions. This is consistent with observations of shock-associated behaviors in other situations (cf. Dunham, et al., 1969). The emotional behaviors generally drop out within a very few sessions of response-contingent punishment training. Of course, this decrement in grid-biting behavior would not be expected in the case of noncontingent punishment in the sense that the UCS is maintained at a particular frequency independent of the changes in behavior.

The data also have implications for the two-process interpretations of punishment suppression. Basically, the very general prediction that some alternative behavior will increase is supported. It is questionable, however, to think that the increase observed in the alternative behaviors observed in the present experiments could account for the suppression in the punished response. The time course of the two transition processes are very different. The punished behavior is suppressed very rapidly relative to the initial disruption and slow rise of the alternative behavior.

The preceding evidence involves the response-contingent punishment operation. The author has not yet started an empirical analysis of the noncontingent punishment procedures. There are, however, several studies in the literature which are relevant to the predictions made by the rules discussed in this article as applied to noncontingent procedures. As indicated in earlier discussion, the organism cannot, by definition, reduce the shock frequency when a noncontingent operation is employed. This leaves the temporal dimension of the contingency in an important role. The response which consistently follows the shock event will predict the absence of the next shock onset for a longer length of time than any other response in the repertoire. The phenomenon reported by Morse et al. (1967) in which an elicited behavior develops and modulates under fixed-interval noncontingent shock suggests that the organism is sensitive to the temporal dimension of the avoidance contingency. More recent evidence reported by Powell and Creer (1969) can be subjected to a similar interpretation. These investigators delivered noncontingent shock at the rate of 100 shocks per session to pairs of rats and recorded the frequency of aggressive attacks per session. The authors did not specify the intershock interval or the session length in their procedure. Ten successive days of noncontingent shock sessions were conducted and measures of aggressive behavior indicated that an increase in the amount of aggression occurred over the course of the experiment. Although the experiment was designed to determine the effects of maturation (among other things) on aggressive behavior, it is suggested that the 10 days of maturation were confounded with 10 days of avoidance training during the experiment. Specifically, the aggressive behavior of these rats predicted the absence of shock onset for a longer period of time than any of the other (unmeasured) behaviors in the organism's repertoire. Similar to the monkeys in the Morse et al. procedure, the rats in the Powell and Creer experiment are assumed to have recognized this avoidance contingency and adjusted the behavior appropriately.

Assuming for the moment that the avoidance interpretation of the Powell and Creer experiment is correct, the present rules have implications for the development of aggression; namely, that certain stimuli will elicit aggressive behaviors, and under a variety of procedures used to deliver the stimuli, the aggressive behavior will develop as an avoidance response and be maintained in the absence of the primary UCS. Under other conditions, the avoidance contingency is not present and one should observe aggression at the unconditioned level. I know of no theoretical account of aggression which implicates an avoidance contingency in the development and maintenance of aggression in animals. Additional work with temporal schedules of aversive stimulation which eliminate this avoidance contingency should reveal the extent to which it is involved in the conditioning of the aggressive behavior of various species.

Some Concluding Remarks

In the absence of more data, it would be premature to attempt to elaborate on the type of theoretical mechanism which is implied by the empirical rules which have been outlined in the preceding discussion. I would like, however, to conclude by suggesting briefly the general directions which one might take in the development of such a mechanism. First, the data and arguments which have been discussed suggest very strongly that there is little to recommend those traditional theoretical accounts of punishment which start from an alternative response assumption. In place of this assumption, the rules which have been outlined imply a symmetrical conditioning mechanism in which (*a*) those responses which predict the aversive event are actively inhibited and (*b*) those responses which predict the absence of the aversive event are actively excited.

History provides at least two examples of symmetrical excitatory and inhibitory mechanisms which deal with appetitive and aversive events. If one is biased toward Pavlovian conditioning mechanisms, it is possible to conceptualize the multiple-response methodology as a *multiple-CS* methodology in which different responses are viewed as different CSs which predict, to varying degrees, the presence or the absence of the unconditioned stimulus (aversive or appetitive). Once conceptualized in terms of Pavlovian operations, some version of Pavlov's (1927) concepts of inhibition and excitation can be developed to explain the effects of punishment on performance. Konorski's (1967) inhibitory and excitatory processes and some recent modifications of these ideas (e.g., Maier, Seligman, & Solomon, 1969) should, for example, be considered in the context of the multiple-response punishment procedure.[4]

---

[4] A basic assumption in Konorski's theorizing is the notion that noxious stimuli such as shock elicit a drive state, fear, which has general inhibitory effects on "preservative" drive states such as hunger, thirst, etc. Punishment would be predicted to have suppressive effect on *all* responses motivated by "preservative" drive states. In the experiments conducted thus far, the inhibitory effects which we have observed are specific to the punished response. For example, when the gerbils were punished for eating they continued to drink the usual amount each session. Hence, if Konorski's inhibition mechanism were to be considered in the context of punishment, the inhibitory properties

Alternatively, if one is biased toward instrumental conditioning mechanisms, it is possible to conceptualize the multiple-response methodology in terms of instrumental contingencies in which each response has either rewarding or punishing consequences. Once conceptualized in terms of instrumental operations, some version of Thorndike's (1913) positive and negative Law of Effect can be elaborated upon as a symmetrical reinforcement mechanism. Rachlin and Herrnstein (1969) have recently made such a suggestion.

In either case, it is hoped that the multiple-response base-line procedure will help to restore the interaction between punishment theory and punishment data which has been lacking in the punishment literature.

---

of shock would have to be more "drive" or "response" specific than he has assumed. Estes (1969, p. 69), in his more recent theorizing, makes an assumption very similar to Konorski's. Specifically, he suggests that the activation of negative drive systems (e.g., shock-produced attack or flight) results in the inhibition of positive drive systems, which, in turn, accounts for the decrement in punished responding. Again, it should be noted that the gerbil data suggest that the only positive drive system which appears to be permanently inhibited during punishment is the specifically punished drive system—or its associated response.

REFERENCES

AZRIN, N. H., & HOLZ, W. C. Punishment. In W. K. Honig (Ed.), *Operant behavior: Areas of research and applications.* New York: Appleton-Century-Crofts, 1966.

BOLLES, R. C. Species-specific defense reactions in avoidance learning. *Psychological Review,* 1970, **77**, 32–48.

BOLLES, R. C., & POPP, R. J., JR. Parameters affecting acquisition of Sidman avoidance. *Journal of the Experimental Analysis of Behavior,* 1964, **7**, 315–321.

DINSMOOR, J. A. Punishment: I. The avoidance hypothesis. *Psychological Review,* 1954, **61**, 34–46.

DINSMOOR, J. A. Punishment: II. An interpretation of empirical findings. *Psychological Review,* 1955, **62**, 96–105.

DUNHAM, P. J., & KILPS, B. Shifts in magnitude of reinforcement: Confounded factors or contrast effects. *Journal of Experimental Psychology,* 1969, **79**, 373–374.

DUNHAM, P. J., MARINER, A., & ADAMS, H. Enhancement of off-key pecking by on-key punishment. *Journal of the Experimental Analysis of Behavior,* 1969, **1**, 156–166.

ESTES, W. K. Outline of a theory of punishment. In B. A. Campbell & R. M. Church (Eds.), *Punishment and aversive behavior.* New York: Appleton-Century-Crofts, 1969.

ESTES, W. K., & SKINNER, B. F. Some quantitative properties of anxiety. *Journal of Experimental Psychology,* 1941, **29**, 390–400.

FALK, J. L. The motivational properties of schedule induced polydipsia. *Journal of the Experimental Analysis of Behavior,* 1966, **9**, 19–25.

KONORSKI, J. *Integrative activity of the brain. An interdisciplinary approach.* Chicago: University of Chicago Press, 1967.

MAIER, S. F., SELIGMAN, M. E. P., & SOLOMON, R. L. Pavlovian fear conditioning and learned helplessness. In B. A. Campbell & R. M. Church (Eds.), *Punishment and aversive behavior.* New York: Appleton-Century-Crofts, 1969.

MILLER, N. E., & DOLLARD, J. *Social learning and imitation.* New Haven: Yale University Press, 1941.

MORSE, W. H., MEAD, R. N., & KELLEHER, R. T. Modulation of elicited behavior by a fixed-interval schedule of electric shock presentation. *Science,* 1967, **157**, 215–217.

MOWRER, O. H. On the dual nature of learning a reinterpretation of "conditioning" and "problem-solving." *Harvard Educational Review,* 1947, **17**, 102–148.

PAVLOV, I. V. *Conditioned reflexes.* (Trans. by G. V. Anrep) New York: Dover, 1927.

POWELL, D. A., & CREER, T. L. Interaction of developmental and environmental variables in shock-elicited aggression. *Journal of Comparative and Physiological Psychology,* 1969, **69**, 219–226.

PREMACK, D. Reinforcement theory. *Nebraska Symposium on Motivation,* 1965, **13**, 123–188.

RACHLIN, H., & HERRNSTEIN, R. J. Hedonism revisited: On the negative law of effect. In B. A. Campbell & R. M. Church (Eds.), *Punishment and aversive behavior.* New York: Appleton-Century-Crofts, 1969.

SCHUSTER, R., & RACHLIN, H. Indifference between punishment and free shock: Evidence for the negative law of effect. *Journal of the Experimental Analysis of Behavior,* 1968, **11**, 777–786.

SOLOMON, R. L. Punishment. *American Psychologist,* 1964, **19**, 239–253.

THORNDIKE, E. L. *Educational psychology.* Vol. 2. *The psychology of learning.* New York: Columbia University, Teacher's College, Bureau of Publications, 1913.

THORNDIKE, E. L. *The fundamentals of learning.* New York: Columbia University, Teacher's College, Bureau of Publications, 1932.

(Received April 21, 1970)

Psychological Review
1971, Vol. 78, No. 1, 71–80

ON THE EXPONENTS IN STEVENS' LAW AND THE CONSTANT IN EKMAN'S LAW[1]

ROBERT TEGHTSOONIAN[2]

Smith College

It follows from Stevens' psychophysical power law, $\psi = \alpha\phi^n$, that the exponent $n = \log R_\psi / \log R_\phi$, where R_ϕ is the ratio of the greatest to the least stimulus intensity and R_ψ is the ratio of corresponding sensory magnitudes. Data from 21 experiments by S. S. Stevens show a correlation (Pearson r) of $-.935$ between $\log R_\phi$ and $1/n$, implying that $\log R_\psi$ is nearly constant. On this basis it is proposed that a single scale of sensory magnitude serves a wide variety of perceptual continua, and that variation in power law exponents is primarily due to variation in dynamic ranges. The hypothesis that there is just one scale of sensory magnitude suggests that there may be just one value for subjective resolving power. When Weber fractions are transformed to their subjective counterparts by the psychophysical power law, the result for nine different continua is nearly constant at about .03.

Since the development of the direct judgment techniques by S. S. Stevens and his associates, an impressive array of data has accumulated showing that judgments of sensory magnitude grow as power functions of stimulus intensity. Each sensory mode yields a characteristic exponent which has been thought to indicate the rate at which log sensory magnitude grows with log stimulus intensity. For example, the exponent of about $\frac{1}{3}$ for brightness has been interpreted as indicative of the decreasing rate at which brightness grows with luminance; at the other extreme, the exponent of 3.5 for judgments of the subjective intensity of electric shock has been regarded as evidence that this experience grows increasingly more rapidly than the current which produces it. Thus, power law exponents can be viewed as measures of an important property of the relevant sensory apparatus.

But other interpretations are possible. Poulton (1967), among others, has raised the possibility that "the sizes of exponents are merely a function of the experimental conditions under which they were determined [p. 316]." Instead of providing us with information about the nervous system, exponents may simply reflect certain parameters of the test situation; the most important of these, according to Poulton, is the geometric range (ratio of extreme intensities) of stimuli employed. As represented in Figure 1, an exponent is the slope of the linear function relating $\log \psi$ (sensory magnitude) to $\log \phi$ (stimulus intensity). To the extent that the subjects' (*S*s') judgments cover a constant range of log magnitudes, exponents will reflect only the range of log intensities selected by the experimenter (*E*). There is then a possibility that *S*s' judgments are to some degree insensitive to the subjective magnitudes which they are supposed to represent, and tend to cover a constant ratio no matter what range of stimuli is presented. In that case, exponents may only be indirect measures of *E*'s arbitrary selection of stimulus intensities. In support of this view, which may be called the "procedural artifact theory," Poulton (1967) reports that for 21 studies conducted by S. S. Stevens and his co-workers, there is indeed a moderate correlation between exponent and geometric range of stimulus intensities (tau = $-.60$).

My purpose in this article is twofold. First, a reexamination of the data considered

---

[1] The manuscript was prepared while the author was a visitor in the Laboratory of Psychophysics, Harvard University. The author is indebted to S. Sternberg, J. C. Stevens, and Martha Teghtsoonian for helpful criticisms of an earlier version of this manuscript.

[2] Requests for reprints should be sent to Robert Teghtsoonian, Department of Psychology, Burton Hall, Smith College, Northampton, Massachusetts 01060.

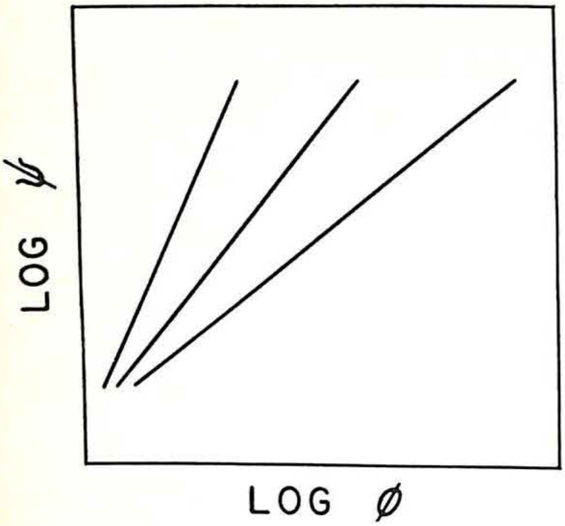


FIG. 1. The effect on power law exponents of variation in range log ϕ when range log ψ is constant.

by Poulton leads me to suggest that he has underestimated the closeness with which exponents are related to the stimulus ranges[3] employed. As is shown below, about 87% of the variance in exponents is accounted for by variation in stimulus range. Second, I wish to propose a hypothesis about this striking relation which is consistent with Stevens' view that exponents reflect an important property of the associated sensory systems. Briefly, this hypothesis proposes that the ratio of the greatest to the smallest possible sensory magnitude is approximately constant for all perceptual continua, and that variation in power law exponents among continua reflects variation in the ratio of the greatest to the smallest stimulus intensity to which *S* is responsive (a ratio which is sometimes called the dynamic range). In short, widely varying dynamic ranges may all be mapped into the same sensory range; and if some nearly constant proportion of those dynamic ranges is presented, the result will still be a nearly constant sensory range, although proportionately smaller than its maximum value.

This hypothesis differs from the procedural artifact theory in two important ways. First, it holds that variation in stimulus range, while certainly under *E*'s control, may be heavily constrained by the sensitivity of *S*'s nervous system. For example, the fact that *E* presents, say, three log units of sound pressure in a loudness scaling experiment but only one log unit of width in scaling the apparent extent of finger span, may reflect the dynamic ranges of the two sensory systems involved rather than *E*'s arbitrariness in determining experimental parameters. Second, this hypothesis holds that the approximate constancy in the ratio spanned by *S*'s judgments is due to the underlying constancy in sensory range rather than to the judgmental rigidity posited by the procedural artifact theory. But before further elaboration of this hypothesis, let us review the evidence considered by Poulton (1967) in assessing the relation between exponents and stimulus range.

S. S. STEVENS' SCALING DATA

These data are drawn from 21 experiments conducted in Stevens' laboratory and are summarized in Poulton's (1967) Table 1. My analysis of these data is based on the following considerations.

It follows from the simple form of the power law

$$\psi = \alpha\phi^n \qquad [1]$$

that

$$n = \frac{\log R_\psi}{\log R_\phi} \qquad [2]$$

where $R\phi$ is the ratio of the greatest to the smallest stimulus intensity, and $R\psi$ is the ratio of corresponding sensory magnitudes. If the range of log judgments provided by *S*s is nearly constant, we may substitute *K* for log $R\psi$. Then

$$n = \frac{K}{\log R_\phi}. \qquad [3]$$

An estimate of this constant *K* may be obtained by determining the value of $n \log R_\phi$ for each of the 21 experiments, and calculating the mean. The resulting value, 1.53, was used in Equation 3 to generate a theoretical function relating the value of the exponent to log R_ϕ. That function is shown in Figure 2 together with the 21 points taken from Poulton's table, each represented by a filled circle. The exponents are plotted on the ordinate against the log stimulus range on the abscissa.[4] (The results of three additional experiments are plotted with unfilled circles. They do not figure in the data analysis, but are included to diminish the importance that a viewer might otherwise attach to the single points at each end of the abscissa.) The quality of the fit can be assessed by replotting the data with the reciprocal of the exponent, $1/n$, on the ordinate against log R_ϕ on the abscissa. If Equation 3 is true, the results should be well fitted by a straight line. Such a fit was made by the method of least squares; a Pearson r of .935 indicates that over 87% of the variance in the reported exponents is accounted for by the stimulus range employed.[5]

---

[3] Throughout this paper, the phrase "stimulus range" refers to the *ratio* of stimulus intensities, and "sensory range" refers to the ratio of corresponding sensory magnitudes.

INTERPRETATIONS

Although the stimulus range may vary from .5 to 6.0 log units, the judgmental range is approximately constant at about 1.5 log units. According to the procedural artifact theory, this constancy represents *S*'s inflexibility in making judgments and, extending Poulton's argument, we could conclude that the exponent is almost entirely determined by *E*'s choice of stimulus range. The alternative interpretation that I have proposed is based on two concepts: (*a*) dynamic range and (*b*) the constancy of the maximum range of subjective magnitude.

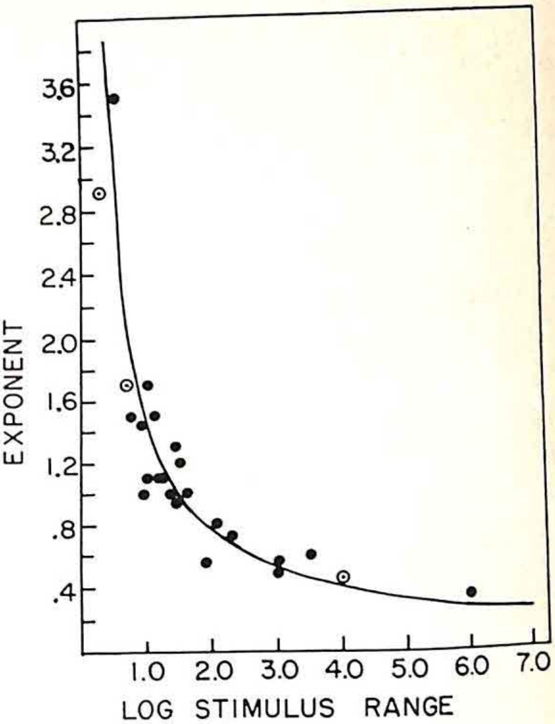


FIG. 2. The filled circles are for 21 experiments by Stevens and his associates tabled by Poulton (1967). The unfilled circles, from left to right, are for saturation of yellow (Indow & Stevens, 1966), saturation of red (Panek & Stevens, 1966), and apparent viscosity (Stevens & Guirao, 1964).

Dynamic Range

The notion that the perceptual continua differ in the ratio of extreme intensities to which *S* can respond is not a novel one, but there is regrettably little direct evidence on its validity. Much of the uncertainty concerns the problem of determining upper bounds to the range of sensitivity. In some cases, there is a fairly obvious physical limitation; for example, finger span simply cannot go much beyond 140 millimeters (mm.), and similarly the capacity to sense a lifted weight will be limited by muscular strength. In other cases, the problem is more difficult, and indeed some may question whether there is an upper limit to sound pressures that can be heard or to luminances that can be seen. Yet even in these cases, the occurrence of physical damage to the receptor system may establish an upper bound.

Despite the obvious difficulties in achieving precise measurement of dynamic range, the concept should not be overlooked as a factor influencing E's choice of stimulus intensities. For example, when scaling the subjective intensity of electric current applied to the fingers, it is just not possible to vary current by much more than a factor of four. By contrast, the ear readily responds to sound pressures varied by a factor of at least 100,000. It is this striking variation in dynamic range that imposes an upper limit on E's selection of R_ϕ, and according to the present hypothesis, it is this same variation which accounts for the variation in power law exponents.

---

[4] The points for binaural loudness, monaural loudness, vocal effort (the autophonic response), and vibration intensity are plotted using values of log R_ϕ which are just half those shown by Poulton. In expressing a given number of decibels as a ratio, Poulton has not taken account of the fact that in these four cases the reported exponents were calculated for ranges of stimulus pressure or amplitude; in such cases the number of decibels is 20 log R_ϕ, not 10 log R_ϕ as Poulton assumed.

[5] Jones and Woskow (1962) report a tau of .93 for nine continua tabled by Stevens (1960). It should be noted that the exponents came from one set of experiments using numerical magnitude estimation, while the stimulus ranges applied to another set of studies using cross-modal matching. Jones and Woskow also were led to surmise that judgmental ranges are approximately constant.

Of course it is unlikely that any of Stevens' 21 experiments made use of the full dynamic range; that is, log R_ϕ/log R_ϕ max was always less than unity. But to the extent that this fraction is nearly the same in all experiments, it would still follow that for log R_ψ a constant, the relative size of exponents reflects the relative size of dynamic ranges. Is log R_ϕ/log R_ϕ max in fact a constant for the 21 studies in question? Lacking definitive measures of dynamic range, a direct empirical answer is not available. But since it is the practice in Stevens' laboratory (as in others) to conduct scaling experiments with the largest value of R_ϕ commensurate with S's comfort, it is not unlikely that a nearly constant proportion of the dynamic range is thereby selected.

Range of Subjective Magnitude

If the maximum value of R_ψ is the same for all perceptual continua, it is of interest to know how large that value may be. The estimate of about 1.5 log units reported above is probably lower than the true value. Suppose, as I have argued above, that log R_ϕ/log R_ϕ max for Stevens' 21 studies is some constant C, less than unity. In that case, an estimate of log R_ψ max is provided by $1.5/C$; and if C is not less than .5, then log R_ψ max is located in the range from 1.5 to 3.0 log units. Only through a determination of values for dynamic range can a more precise estimate be developed.[6]

What kind of perceptual structure might generate the same value of R_ψ max for all continua? The present data are consistent with a model that posits a single central mechanism responsible for all judgments of sensory magnitude. Such a sensory magnitude monitor would impose the same range of possible outputs regardless of the receptor system in the periphery to which it was responding. In this context, the various receptor systems can be regarded as performing the necessary expansions or compressions required to map the widely varying dynamic ranges into this constant range of subjective magnitudes. Empirical investigation is required for a more rigorous evaluation of this construct and the development of additional defining properties. For the present, it is proposed only for its heuristic value in conceptualizing a structural basis for the constant in Equation 3.

INTRAMODAL VERSUS INTERMODAL VARIATION IN STIMULUS RANGE

Given the two interpretations of Equation 3, are there grounds for preferring one to another? Do they generate differential predictions? The present hypothesis of a constant sensory range relates exponents to an inherent characteristic of the relevant receptor system and does not in itself anticipate any effect of intradmodal variation in R_ϕ. But as the procedural artifact theory regards the actual value of R_ϕ as critical, the effect on exponents should be no different whether R_ϕ is varied intermodally (as in Figure 2) or intramodally. Indeed, Poulton's (1968) review of the pertinent literature reveals that for some continua there are conditions under which the exponent declines as R_ϕ grows. For example, Stevens and Poulton (1956) had separate groups of Ss make magnitude estimations of a 1000-Hertz (Hz.) tone 6, 10, 20, or 40 decibels (db.) less than a standard set at about 100 db. above threshold. The exponents estimated for these conditions are approximately .67, .60, .52, and .46, respectively. However, if these values are referred to the coordinates of Figure 2, it is evident by inspection that this effect of variation in log R_ϕ falls far short of an inverse relation to n. It seems that loudness judgments cannot be made to grow as rapidly as judgments of the subjective intensity of, say, electric shock by using the same value of R_ϕ for sound pressure as for electric current. While more evidence is needed to resolve this point, the results of Stevens and Poulton suggest that the intramodal effects of variation in R_ϕ may be quite different in magnitude, and hence probably in cause, from the intermodal effects depicted in Figure 2.

---

[6] An estimate of log R_ψ can also depend on the exact form of the power law which is employed. The version which translates stimulus intensity by an additive constant, $\psi = \alpha(\phi - \phi_0)^n$, can be used in conjunction with the present data to generate an estimate of log $R_\psi = 2.11$.

RESOLVING POWER

What has been called Ekman's law (Stevens, 1966) states that the subjective size of the just noticeable difference (jnd) is linearly proportional to the subjective magnitude of the standard; there appears to be a relativity for sensory magnitudes analogous to that defined by Weber's law for stimulus magnitudes. But if there is just one range of sensory magnitudes, it is tempting to consider that there may be just one constant of proportionality for Ekman's law. Thus S might be able to resolve a 100 $C\%$ change in sensory magnitude (where $C = \Delta\psi/\psi$) regardless of the form of stimulus input. Such a conclusion is consistent with the hypothesis that discriminations are based on the output of the sensory magnitude monitor discussed above.

If one selects the simplest version of the power law,

$$\psi = \alpha\phi^n,$$

then it follows that, for any continuum,

$$\psi + \Delta\psi = \alpha(\phi + \Delta\phi)^n$$

and therefore

$$1 + \frac{\Delta\psi}{\psi} = \left(1 + \frac{\Delta\phi}{\phi}\right)^n \qquad [4]$$

where $\Delta\phi/\phi$ is the Weber fraction and $\Delta\psi/\psi$ is the associated fractional growth in ψ. To test the constancy of $\Delta\psi/\psi$ requires that empirical values of $\Delta\phi/\phi$ obtained for various continua be inserted in Equation 4. An unequivocal specification of $\Delta\phi/\phi$ is notoriously difficult; the value is dependent on many parameters of the test situation. Nonetheless, it seemed reasonable to search for values of $\Delta\phi/\phi$ that could be regarded as characteristic of a particular continuum. The values of $\Delta\phi/\phi$ shown in Table 1 are for stimuli that resemble as closely as possible the stimuli that were employed in the scaling experiments which generated the tabled exponents. In the following sections, the sources for these data are cited, and any analyses which go beyond what the authors provide are explained.

TABLE 1

INVARIANCE OF $[1 + (\Delta\phi/\phi)]^n = 1 + (\Delta\psi/\psi)$

Continuum	$\Delta\phi/\phi$	n	$\Delta\psi/\psi$
Brightness	.079	.33	.026
Loudness	.048	.6	.029
Finger span	.022	1.30	.029
Heaviness	.020	1.45	.029
Length	.029	1.04	.030
Taste, NaCl	.083	.41	.033
Saturation, red	.019	1.7	.033
Electric shock	.013	2.5	.033
Vibration			
60 Hz.	.036	.95	.034
125 Hz.	.046	.67	.031
250 Hz.	.046	.64	.029
Mean			.031
Grand mean for nine continua			.030

Note.—$\Delta\phi/\phi$ is the Weber fraction, $\Delta\psi/\psi$ is the associated fractional growth in sensory magnitude, and n is the exponent of the power function relating subjective magnitude to physical intensity.

The Data

Brightness. One must be especially uneasy about selecting a single value to represent the Weber constant for brightness discrimination, since measures of resolving power depend at least on target size, duration, and retinal location, as well as the state of the retina at the time of the test. But there is also evidence that the measured size of the difference limen (DL) becomes increasingly insensitive to changes in duration and area beyond certain critical levels. For duration, Anglin and Mansfield (1968) have summarized evidence that at very high luminances (around 95 db.) 25 milliseconds is the critical value; longer exposures will

TABLE 2

SUMMARY OF FIVE STUDIES OF DIFFERENTIAL SENSITIVITY FOR BRIGHTNESS

Study	Method	No. *S*s	Target diameter (minutes)	Duration (seconds)	Log $\Delta\phi/\phi$
Herrick (1956)	Limits, ascending	2	60	.032 to 2.013	−1.2
Hattwick (1954)	Limits, ascending	1	60	.02	−1.1
Mueller (1951)	Constant stimuli 60% point	2	40	.02	−1.1
Keller (1941)	Limits, ascending and descending	2	88	.20	− .9
Graham & Bartlett (1940)	Limits, ascending and descending	2	56	.03	−1.4

not further reduce the DL. In the studies to be considered, exposure times were near or greater than this critical duration. For target size, relatively little improvement in differential sensitivity occurs beyond diameters of about 1° (e.g., Blackwell, 1946); the studies to be considered all employed targets near that size. To obtain from those studies the best estimate of the slope constant in the linear generalization of Weber's law, the value of the Weber fraction was taken at the highest intensity studied (or the point at which the fraction was minimized). Five experiments which met these conditions are summarized in Table 2; they are in fairly good agreement about the value of log $\Delta\phi/\phi$ despite variation in psychophysical method and experimental conditions. The median value of log $\Delta\phi/\phi$, -1.1, was selected and its antilog is the value shown in Table 1. Of course the list of studies in Table 2 is not exhaustive, but the reported values seem to be representative. Two well-known studies which report markedly lower values of $\Delta\phi/\phi$, Steinhardt (1936) and Blackwell (1946), were excluded because of the unusual degree of training provided the *S*s; both authors comment on the many thousands of judgments made over prolonged periods of time. Further, Blackwell's *S*s were encouraged (successfully) to respond at levels of awareness well below those usually required by the classical psychophysical methods.

The exponent of the power law describing the brightness function has been reported as .33 by Stevens and Stevens (1963).

Loudness. For loudness, the value .048 is due to Miller (1947) and refers to discrimination of the intensity of white noise. He found that at levels greater than 30 db. above threshold, the increment (1.5-second pulse) heard 50% of the time was nearly constant at .41 db.; his sample was composed of two experienced *S*s. A similar value has been reported by Harris (1963) for a wide range of pure tones (125 Hz. to 6000 Hz.) using the method of constant stimuli. The exponent of .6 for the power function relating loudness to sound intensity was reported by Stevens (1956) and has been recognized as the standard value.

Finger span. Differential sensitivity for finger span has been studied by Gaydos (1958) using the method of adjustment. In his Experiments 1 and 2, a total of 100 *S*s made matches to standards ranging from 17.7 to 100 mm. Probable errors can be plotted against the lengths of the standards, yielding a linear function with a slope of .024. In his Experiment 3, Gaydos reports data obtained from a group of 20 *S*s for a shorter range of standards; ignoring an anomalous value for the shortest standard, a linear fit of the probable errors to the standards yields a slope of .020. A similar value, .019, has been reported by Stevens and Stone (1959) for data obtained by G. J. Huberman.[7] Giving double weight to the value derived from Gaydos' Experiments 1 and 2, these three results have a mean of .022.

Stevens and Stone (1959) reported that magnitude estimation of apparent finger span yields a power function with an exponent of 1.33. More recently, Mashhour and Hosman (1968) have obtained an exponent of 1.26 using a standard in the middle of a series of 11 stimuli ranging from 10 mm. to 110 mm. The mean of these two results, rounded to 1.30, is the value listed in Table 1.

---

[7] Stevens and Stone (1959) plotted standard deviations rather than probable errors and reported a slope of .0286. I have multiplied this by .6745 to obtain the slope for probable errors.

Heaviness. For differential sensitivity to lifted weights, the work of Oberlin (1936) is most frequently cited. In his Experiment II, five *S*s judged the second of pairs of weights as heavier, lighter, or equal in comparison to the first. Standards varied from 25 grams (gm.) to 600 gm. Urban's process was used to determine the upper and lower DLs. When half of the interval is calculated, the average over *S*s is nearly linear with the standard weight (although the value for 600 gm. is aberrant); the slope is .02. This slope fits the data of Experiment I (in which only "heavier" and "lighter" judgments were allowed) over the lowest four standards, 50 gm. to 250 gm., for each of four *S*s.

The exponent of the power function for apparent heaviness is 1.45 as reported by Stevens and Galanter (1957). Although Mashhour and Hosman (1968) reported 1.13 for a range from 70 gm. to 700 gm. with a 222-gm. standard, the figure reported by Stevens and Galanter is an average over several studies and is therefore preferred.

Length. In measuring DLs for length, Ono (1967) employed a variant of the method of limits for standard lengths of 5, 10, and 15 centimeters, at distances of 1.5, 3.0, and 4.5 meters. Although standard and comparison stimuli were displayed together, their lateral displacement was sufficient to ensure nonsimultaneity of viewing. When the size of the DL is plotted against the distal size of the standard, the slope of the linear relation is .029. Stevens and Galanter (1957) reported that magnitude estimations of apparent length are "very nearly a linear function of physical length"; the exponent of the power function is close to 1.00. Mashhour and Hosman (1968) reported an exponent of 1.08 for lengths ranging from 15 mm. to 150 mm. with a standard of 55 mm. The mean of these two values is 1.04.

Saltiness. McBurney, Kasschau, and Bogart (1967) measured the size of the jnd for NaCl after adaptation to water, and also after adaptation to a .1 molar solution of NaCl. The results of the latter condition, although representing just one point on the function relating the Weber fraction to intensity, are reported here. The average of $\Delta\phi/\phi$ for three *S*s following a method of constant stimuli was .09. Using an ascending method of limits with a criterion of two successive correct responses, the average $\Delta\phi/\phi$ for two *S*s was .075. The mean of these two values, .083, is used. McBurney (1966) used magnitude estimation with 10 *S*s judging the apparent intensity of solutions of NaCl ranging from water and .001 molar to 1.0 molar concentrations. Using empirically determined threshold corrections for each of several adapting intensities, the data can be well described by a power function with an exponent of .41.

Saturation. Panek and Stevens (1966) obtained magnitude estimations of the saturation of red, and measured jnd over a range of saturations. The latter was done using the method of constant stimuli for four *S*s and the method of single stimuli for four *S*s, for saturations varying from 20% to 80% red. The jnd was defined as *Q*, half the distance between the 25% and 75% values, and the results averaged over both methods are reported. The relation between *Q* and the standard saturation is well fitted by a straight line (discounting the value for the weakest standard) with a slope of .019. Data are also reported for the method of constant stimuli alone at standards of 20% and 80%; a straight line connecting these two values of *Q* has a slope of .019, and this is the value shown in Table 1. The results of the several magnitude estimation experiments are well described by a power function with an exponent of 1.7.

Shock. Hawkes (1961) measured Weber fractions for the intensity of electric shock to the finger tip, for a range of frequencies, using both the beat method and the method of successive stimuli. From his graphical display of Weber fractions, values of $\Delta\phi$ were calculated for each of three *S*s. Only two intensities were studied, 120% and 200% of each *S*'s absolute threshold, so it is the slope of the straight line connecting these two points in the plot of $\Delta\phi$ against ϕ that can be calculated. Since there appears to be little effect due to psychophysical method or frequency, slopes were obtained for all three *S*s at 100 Hz. and at 1500 Hz. using both methods. The median value of this (somewhat skewed) distribution of 12 scores is .013.

Stevens, Carton, and Shickman (1958) scaled the subjective intensity of electric shock covering a range from .38 milliampere (ma.) to 1.15 ma. and reported a power function with an exponent of 3.5. Ekman, Frankenhaeuser, Levander, and Mellis (1966) had 15 *S*s judge the unpleasantness of electric shock varying in intensity from 1.5 to 3.5 times each *S*'s mean sensation threshold. The results together with findings from a previous study by the same authors are well fitted by Ekman's version of the power law (involving an additive translation of stimulus intensity) with an exponent of 1.54. The mean of these two values is 2.5.

Vibration. The case of tactile vibration is interesting because the exponent of the power function varies with frequency. Stevens (1959) has reported that magnitude estimation of a 60-Hz. vibration to the fingertip follows a power function with an exponent of .95. Later, Stevens (1968) found that if a 60-Hz. vibration is adjusted to match a 125-Hz. vibration or a 250-Hz. vibration, the resulting power functions have exponents of .70 and .67, respectively. These two values may be multiplied by .95 to generate estimates of the exponents—.67 and .64—of power functions obtained by magnitude estimation.

Knudsen (1928) measured DLs for vibration at 64 Hz., 128 Hz., and 256 Hz., values which are similar to those employed in Stevens' scaling studies. He employed a method of limits with both ascending and descending series. His data are average values for two *S*s and are presented in the form of Weber fractions. From his graphs, estimates of $\Delta\phi$ can be derived, and when plotted against ϕ (expressed relative to approximate threshold), are well fitted by straight lines. The slopes are .036, .046, and .046 for vibration frequencies of 64 Hz., 128 Hz., and 256 Hz., respectively.

Invariance of $\Delta\psi/\psi$

The apparent stability in the computed values of $\Delta\psi/\psi$ is impressive, especially in view of the difficulty of demonstrating, first, that the several values of $\Delta\phi/\phi$ are comparable, and second, that they are each preferable to alternatives that may be found in the literature. Despite these difficulties, the values of Table 1 raise the possibility that a constant percentage change in sensory magnitude is just discriminable, regardless of the nature of the stimulus. The view that the steepness of the psychophysical function is directly related to differential sensitivity was of course Fechner's idea, and has in recent years provoked severe criticism (e.g., Stevens, 1961), as well as frequent restatements (e.g., Heinemann, 1961; Luce, 1959). But it is important to recognize, as Heinemann (1961) has pointed out, that this aspect of Fechner's speculation is quite separable from his ill-fated hypothesis that jnd's define subjectively constant *intervals*. Instead, what seems to be constant is the *ratio* of subjective magnitudes generated by stimuli one jnd apart, and the evidence of Table 1 shows that the value of this ratio is constant for a wide variety of continua.

A final word of caution may be useful. A critic could of course dispute any one of the empirical values in Table 1; the Weber constants are especially vulnerable, and there are undoubtedly anomalous cases that do not fit the pattern. But it seems evident that, without being subject to gross distortion, many of the data in the discrimination literature fall into a simple and conceptually attractive pattern. What remains to be seen is whether this pattern extends over an even broader domain than is indicated here, or whether it is liable to so many exceptions that its value is lost.

In summary, one can point to two large areas of data suggesting that the experience of sensory magnitude is governed by a single mechanism. Two of its properties have been described here; further research may reveal others.

CONCLUSIONS

Three hypotheses have been proposed together with supporting evidence:

1. There is a common scale of sensory magnitude for a wide variety of perceptual continua.

2. The maximum range of log sensory magnitudes defined by this scale accounts for the inverse relation between power law exponents and the range of log stimulus intensities, when the latter are proportional to their maximum values (dynamic ranges).

3. Data from a number of discrimination studies are consistent with the view that just noticeable changes occur when sensory magnitudes are altered by a constant fraction, and that this fraction is the same regardless of the form of input energy.

REFERENCES

ANGLIN, J. M., & MANSFIELD, R. J. W. On the brightness of short and long flashes. *Perception and Psychophysics*, 1968, **4**, 161–162.

BLACKWELL, H. R. Contrast thresholds of the human eye. *Journal of the Optical Society of America*, 1946, **36**, 624–643.

EKMAN, G., FRANKENHAEUSER, M., LEVANDER, S., & MELLIS, I. The influence of intensity and duration of electrical stimulation on subjective variables. *Scandinavian Journal of Psychology*, 1966, **7**, 58–64.

GAYDOS, H. F. Sensitivity in the judgment of size by finger-span. *American Journal of Psychology*, 1958, **71**, 557–562.

GRAHAM, C. H., & BARTLETT, N. R. The relation of the size of the stimulus and intensity in the human eye: III. The influence of area on foveal intensity discrimination. *Journal of Experimental Psychology*, 1940, **27**, 149–159.

HARRIS, J. D. Loudness and discrimination. *Journal of Speech and Hearing Disorders*, 1963 (Monogr. Suppl. 11).

HATTWICK, R. G. Dark adaptation to intermediate levels into complete darkness. *Journal of the Optical Society of America*, 1954, **44**, 223–228.

HAWKES, G. R. Cutaneous discrimination of electrical intensity. *American Journal of Psychology*, 1961, **74**, 45–53.

HEINEMANN, E. G. The relation of apparent brightness to the threshold for differences in luminance. *Journal of Experimental Psychology*, 1961, **61**, 389–399.

HERRICK, R. M. Foveal luminance discrimination as a function of the duration of the increment in luminance. *Journal of Comparative and Physiological Psychology*, 1956, **49**, 437–443.

INDOW, T., & STEVENS, S. S. Scaling of saturation and hue. *Perception and Psychophysics*, 1966, **1**, 253–272.

JONES, F. N., & WOSKOW, M. J. On the relationship between estimates of magnitude of loudness and pitch. *American Journal of Psychology*, 1962, **75**, 669–671.

KELLER, M. The relation between the critical duration and intensity in brightness discrimination. *Journal of Experimental Psychology*, 1941, **28**, 407–418.

KNUDSEN, V. O. "Hearing" with the sense of touch. *Journal of General Psychology*, 1928, **1**, 320–352.

LUCE, R. D. *Individual choice behavior.* New York: Wiley, 1959.

MASHHOUR, M., & HOSMAN, J. On the new "psychophysical law": A validation study. *Perception and Psychophysics*, 1968, **3**, 367–375.

MCBURNEY, D. H. Magnitude estimation of the taste of sodium chloride after adaptation to sodium chloride. *Journal of Experimental Psychology*, 1966, **72**, 869–873.

MCBURNEY, D. H., KASSCHAU, R. A., & BOGART, L. M. The effect of adaptation on taste jnds. *Perception and Psychophysics*, 1967, **2**, 175–178.

MILLER, G. A. Sensitivity to changes in the intensity of white noise and its relation to masking and loudness. *Journal of the Acoustical Society of America*, 1947, **19**, 609–619.

MUELLER, C. G. Frequency of seeing functions for intensity discrimination. *Journal of General Physiology*, 1951, **34**, 463–474.

OBERLIN, K. W. Variation in intensive sensitivity to lifted weights. *Journal of Experimental Psychology*, 1936, **19**, 438–455.

ONO, H. Difference threshold for stimulus length under simultaneous and nonsimultaneous viewing conditions. *Perception and Psychophysics*, 1967, **2**, 201–207.

PANEK, D. W., & STEVENS, S. S. Saturation of red: A prothetic continuum. *Perception and Psychophysics*, 1966, **1**, 59–66.

POULTON, E. C. Population norms of top sensory magnitudes and S. S. Stevens' exponents. *Perception and Psychophysics*, 1967, **2**, 312–316.

POULTON, E. C. The new psychophysics: Six models for magnitude estimation. *Psychological Bulletin*, 1968, **69**, 1–19.

STEINHARDT, J. Intensity discrimination in the human eye: I. The relation of ΔI/I to intensity. *Journal of General Physiology*, 1936, **20**, 185–209.

STEVENS, J. C., & STEVENS, S. S. Brightness function: Effects of adaptation. *Journal of the Optical Society of America*, 1963, **53**, 375–385.

STEVENS, S. S. The direct estimation of sensory magnitudes—loudness. *American Journal of Psychology*, 1956, **69**, 1–25.

STEVENS, S. S. Tactile vibration: Dynamics of sensory intensity. *Journal of Experimental Psychology*, 1959, **57**, 210–218.

STEVENS, S. S. The psychophysics of sensory function. *American Scientist*, 1960, **48**, 226–253.

STEVENS, S. S. To honor Fechner and repeal his law. *Science*, 1961, **133**, 80–86.

STEVENS, S. S. A metric for the social consensus. *Science*, 1966, **151**, 530–541.

STEVENS, S. S. Tactile vibration: Change of exponent with frequency. *Perception and Psychophysics*, 1968, **3**, 223–228.

STEVENS, S. S., CARTON, A. S., & SHICKMAN, G. M. A scale of apparent intensity of electric shock. *Journal of Experimental Psychology*, 1958, **56**, 328–334.

STEVENS, S. S., & GALANTER, E. H. Ratio scales and category scales for a dozen perceptual continua. *Journal of Experimental Psychology*, 1957, **54**, 377–411.

STEVENS, S. S., & GUIRAO, M. Scaling of apparent viscosity. *Science*, 1964, **144**, 1157–1158.

STEVENS, S. S., & POULTON, E. C. The estimation of loudness by unpracticed observers. *Journal of Experimental Psychology*, 1956, **51**, 71–78.

STEVENS, S. S., & STONE, G. Finger span: Ratio scale, category scale, and JND scale. *Journal of Experimental Psychology*, 1959, **57**, 91–95.

(Received May 11, 1970)

Psychological Review
1971, Vol. 78, No. 1, 81-82

LETTERS TO THE EDITOR

ON ACQUIESCENCE AND SOCIAL DESIRABILITY

In a recent article in this Journal, Dillehay (Dillehay, R. C. Sincerity and dogmatism: A reassessment and new data. *Psychological Review*, 1969, **76**, 422–424) criticizes an article by Stanley and Martin (Stanley, G., & Martin, J. How sincere is the dogmatist? *Psychological Review*, 1964, **71**, 331–334) as not providing a proper test of Rokeach's Hypothesis B_1 (Rokeach, M. The double agreement phenomenon: Three hypotheses. *Psychological Review*, 1963, **70**, 304–309). This hypothesis asserts (*a*) that both original and reversed items on the F and D scales are high on social desirability; (*b*) that high scorers on the original scales disagree with reversed items; but that (*c*) their response to reversed items is based on the social desirability dimension, resulting in scores apparently reflecting low scores on F and D.

Stanley and Martin correctly deduce from this hypothesis that a positive correlation should exist between acquiescent subjects' responses to reversed items on the D scale and their scores on a social desirability scale. Unfortunately, they do not make clear either in the summary or background that the correlation is deduced to exist for acquiescent respondents only, and Dillehay's main criticism springs from his understandable failure to appreciate this.

From Stanley and Martin's failure to find such a correlation, it does follow that they have "no evidence to support Rokeach's Hypothesis B_1 [p. 333]." True, as Dillehay states, they do not "refute" the hypothesis, but this is for the *logical* reason that failure to find confirmation is not refutation.

Dillehay correctly deduces from Rokeach's hypothesis that acquiescent respondents are higher on social desirability than those who are consistently high on dogmatism. His findings support this deduction, but it is weak confirmation, since there is no need to assume active disagreement by acquiescent respondents: uncertainty would be enough.

Each of Rokeach's three hypotheses may account for some acquiescence, but Hypothesis B_1 cannot account for all acquiescence, since factors such as the respondents' uncertainty are important (Bethlehem, D. W. Dogmatism and allied personality variables related to prejudice in young adults. Unpublished doctoral dissertation, University of London, 1969).

(Received September 21, 1970)

Douglas W. Bethlehem
Subject of Psychology
University of Zambia
P.O. Box 2379
Lusaka, Zambia

SUBJECTIVE TIME AND DECISION MAKING

Pollay (Pollay, R. W. A model of decision times in difficult decision situations. *Psychological Review*, 1970, 77, 274–281) assumes that "most decision makers . . . will continue to expend effort in the analysis of a decision problem only as long as the expected returns from that effort exceed the costs of engaging in the analysis [p. 274]." It seems more rational to assume that they "will continue . . . as long as the expected returns from that effort exceed the *expected* costs. . . ." However, the expected costs change with each elimination of an alternative. Furthermore, the expected returns may be inversely related to the expected costs, if the decision maker finds the decision-making process in itself gratifying because he enjoys its challenges. This return may change with time as the decision maker becomes either more fascinated or more bored with the decision-making process.

Thus, time must be considered in terms of subjective time for each decision maker and not in terms of chronometer time which cannot account for the subjective cost to the decision maker. Subjective time, however, probably does not produce the linear function assumed by Pollay in his first "subsidiary assumption." It is also not merely a function of the decision maker's personality but of the latter in relation to the total situation (including what else the decision maker may want to do in the near future besides deciding among the alternatives presented to him in the decision-making situation).

True indeterminacy is evident in the problem Pollay attempts to solve when his model is seen as an open system (Berrien, F. K. *General and social systems.* New Brunswick, N. J.: Rutgers University, 1968) in which the decision to "quit or give up" is considered as a decision within a suprasystem in which the continuation of the experimental decision-making situation is only one of several alternatives.

(Received October 9, 1970)

Lewis W. Brandt
Department of Psychology
Division of Social Sciences
University of Saskatchewan
Regina, Canada

The British Journal of Psychology

Volume 61 / Part 4 / November 1970

E. STURE ERIKSSON
A field theory of visual illusions

M. T. TURVEY, P. BRICK, AND J. OSBORN
Temporal course of proactive interference in short-term memory

G. R. J. HOCKEY
Changes in attention allocation in a multicomponent task under loss of sleep

JOHN C. WEBSTER, MURIEL M. WOODHEAD, AND ALAN CARPENTER
Perceptual constancy in complex sound identification

R. A. KENNEDY AND A. L. WILKES
A comparison of retrieval and location operations in active and passive sentences

JOHN L. BRADSHAW
Phonetic homogeneity and articulatory lengthening

P. C. WASON, AND P. N. JOHNSON-LAIRD
A conflict between selecting and evaluating information in an inferential task

ALISTAIR E. PHILIP
Reliability and equivalence in the eight-parallel form anxiety battery

RICHARD H. LENT
Binocular resolution and perception of race in the United States

S. A. MAGOWAN AND TERENCE LEE
Some sources of error in the use of the projective method for the measurement of moral judgement

MARIAN ANNETT
The growth of manual preference and speed

H. R. SCHAFFER AND M. H. PARRY
The effects of short-term familiarization on infants' perceptual-motor co-ordination in a simultaneous discrimination situation

Annual subscription $25.00 for four issues

CAMBRIDGE UNIVERSITY PRESS
32 East 57th Street, New York, N. Y. 10022

9 DEC 1971

Psychology

ABILITIES: THEIR STRUCTURE, GROWTH, AND ACTION
Raymond B. Cattell, University of Illinois

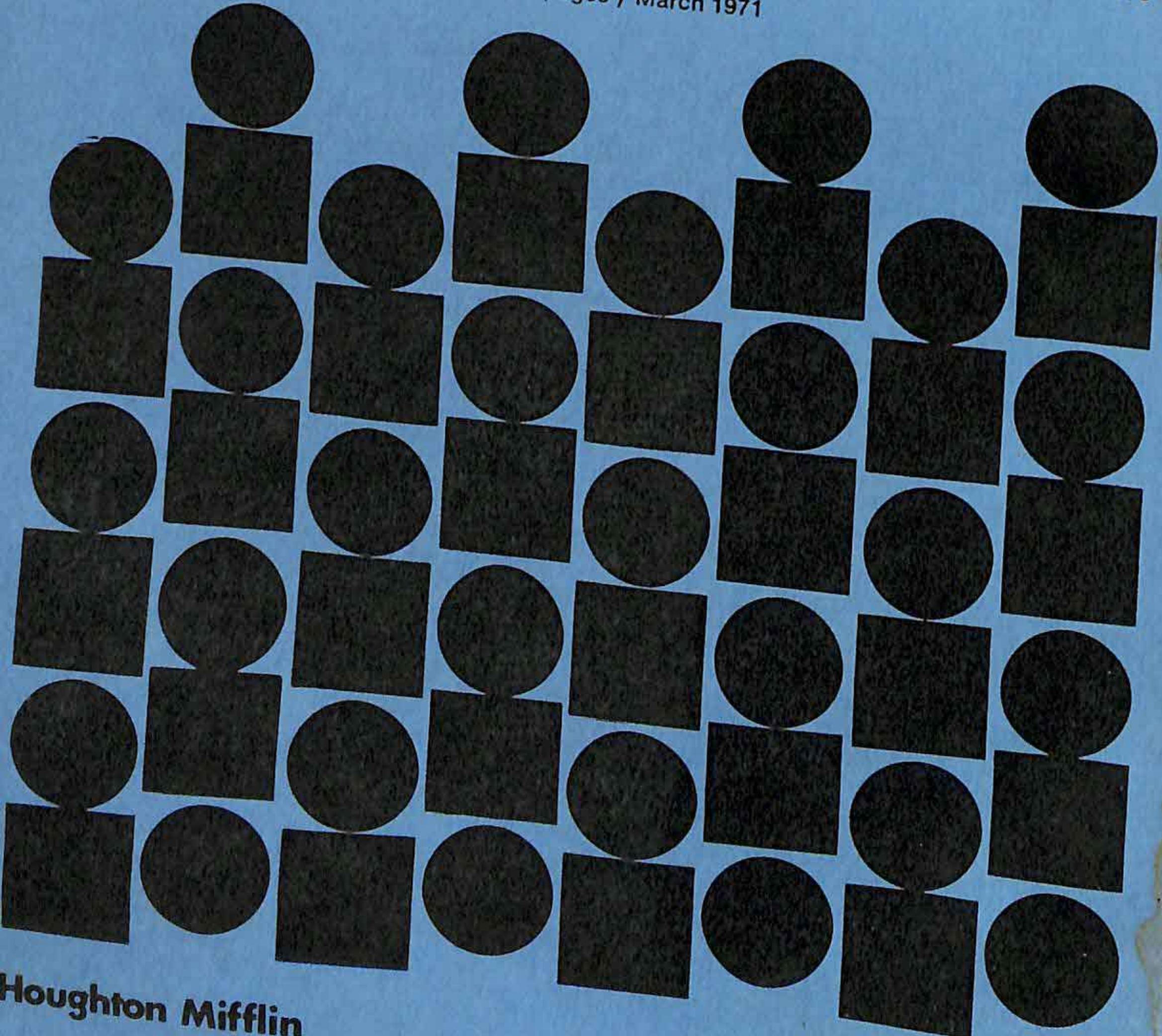
Dr. Cattell examines the field of abilities from the perspective of personality and motivation research and integrates it with physiological and neurological psychology, comparative animal psychology, machine simulation, and developmental psychology.
About 780 pages / June 1971

MIND IN EVOLUTION: An East-West Synthesis of Learned Behavior and Cognition
Gregory Razran, Queens College

MIND IN EVOLUTION is an evolutionary synthesis of the main categories of systematic psychology: Russian Pavlovianism, American behaviorism, and Western cognitionism. Dr. Razran has unified the once seemingly separate evidence and logic of these major schools into one hierarchical system using the framework of Darwinian evolution.
About 440 pages / March 1971

PERSONALITY DEVELOPMENT: From Infancy to Adulthood
Charles Wenar, The Ohio State University

An eclectic, up-to-date approach to understanding the development of personality variables which determine normal and deviant personality, PERSONALITY DEVELOPMENT presents in a topically organized discussion a description of development and a theoretical accounting for the descriptive data.
454 pages / December 1970



Houghton Mifflin

Publisher of The American Heritage Dictionary of the English Language
Boston / Atlanta / Dallas / Geneva, Ill. / New York / Palo Alto

21.9.71

Ke
Checked & to-
Issue the librarian
R Dean
26/10/71

PSYCHOLOGICAL REVIEW

GEORGE MANDLER, EDITOR
JEAN M. MANDLER, ASSOCIATE EDITOR
UNIVERSITY OF CALIFORNIA, SAN DIEGO

Vol. 78, No. 2

March 1971

Articles

Limbic-Diencephalic Mechanisms of Voluntary Movement . . . . **C. H. Vanderwolf** 83

Visual Feature-Analyzers and Aftereffects of Tilt and Curvature . . . . **Max Coltheart** 114

Problem of Serial Order in Behavior Is Not Resolved by Context-Sensitive Associative Memory Models **Terry Halwes and James J. Jenkins** 122

Hypothesis Theory and Nonlearning Despite Ideal S–R-Reinforcement Contingencies . . . . **Marvin Levine** 130

An Analysis of Two Quantitative Theories of Cognitive Balance **A. Rodney Wellens and Donald L. Thistlethwaite** 141

Conjoint-Measurement Analysis of Composition Rules in Psychology . . . . . **David H. Krantz and Amos Tversky** 151

Letters to the Editor 170

Published Bimonthly
by The American Psychological Association, Inc.

Editorial Consultants

Robert P. Abelson
Bernard Agronoff
Norman H. Anderson
Samuel Barondes
Jacob Beck
Thomas G. Bever
R. L. Birdwhistell
Donald S. Blough
Lyle E. Bourne, Jr.
Gordon Bower
Martin D. Braine
Donald T. Campbell
Barry E. Collins
Stanley Coopersmith
C. G. Costello
Phebe Cramer
Paul Ekman
Norman T. Feather
Martin E. Fishbein
Rochel Gelman
Kenneth Gergen
Henry Gleitman
Stephen Glickman
James Greeno
S. P. Grossman
C. S. Harris
Samuel Himmelfarb
James R. Ison
Charles Kiesler
Paul A. Kolers
Peter J. Lang
A. M. Liberman
R. Duncan Luce
Dominic Massaro
David C. McClelland
David McNeill
Ulric Neisser
Clyde E. Noble
Dean Peabody
W. F. Prokasy
Frank Restle
David E. Rumelhart
Irwin G. Sarason
T. Shallice
George Sperling
Sally Sperling
Saul Sternberg
Lise Wallach
William Wiest
Robert S. Wyer

(A large group of psychologists has kindly agreed to assist in the editing of manuscripts submitted to the *Psychological Review*. In each issue the Editor will list in the *Review* those editorial consultants who have helped him process manuscripts for that issue, so that their efforts may be gratefully acknowledged.)

The *Psychological Review* is devoted to articles of theoretical significance to any area of scientific endeavor in psychology. Except for occasional articles solicited by the Editor, manuscripts exceeding 7,500 words will not be accepted. Ordinarily, manuscripts which consist primarily of original reports of research or of reviews of the literature should be submitted to other journals. Methodological and statistical articles are not appropriate for the *Psychological Review*. See the editorials in the *Review* (1965, **72**, 1–2; 1971, **78**, 1–2) for further details.

Beginning with the 1971 issue, Letters to the Editor will be published on an experimental basis. Letters not exceeding 250 words on topics of theoretical interest should be submitted in duplicate.

Manuscripts must be accompanied by an abstract of 100 to 120 words typed on a separate sheet of paper. The abstract should conform to the style of *Psychological Abstracts*. Detailed instructions for preparation of the abstracts appeared in the *American Psychologist* (1961, **16**, 833), or may be obtained from the APA Central Office.

All manuscripts must be submitted in triplicate. In the reference list, give journal titles in full; do not abbreviate. Figures should be submitted as glossy prints; duplicate figures may be photographic or pencil-drawn copies. Authors should retain a copy of the manuscript in case there is loss in the mail.

Because of the large number of manuscripts submitted, there is an inevitable publication lag of several months. Authors may avoid this delay if they are prepared to pay the costs of publishing their own articles; the appearance of articles by other contributors is not thereby delayed.

Manuscripts should be addressed to the Editor, George Mandler, Department of Psychology, University of California, San Diego, La Jolla, California 92037.

Copyrighted material. The following policies govern reprinting or other forms of reproduction of materials copyrighted by the American Psychological Association: Written permission must be obtained for copying or reprinting any tables, figures, or text of more than 500 words in length. Persons or organizations who obtain permission to reproduce any such copyrighted materials must include on the first page of the reproduced materials the exact copyright notice which appeared on the Association's original publication. There will be a charge of $10.00 per page for such permission except when the requester is the author of the article to be reprinted, or when the article is to be reproduced in a limited number of copies solely for instructional purposes.

Helen Orr
Managing Editor

Lorraine Biros
Technical
Editor

Henry S. Salmon
Advertising and Business
Manager

Robert D. Gorman
Subscription Manager

Published Bimonthly by the
AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
Prince and Lemon Streets, Lancaster, Pa.
1200 Seventeenth St., N. W., Washington, D. C. 20036

$10.00 volume

$2.00 issue

Second-class postage paid at Lancaster, Pa.

Send all communications, including address changes, to 1200 Seventeenth St., N. W., Washington, D. C. 20036. Address changes must arrive by the tenth of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

PSYCHOLOGICAL
REVIEW

VOL. 78, NO. 2

MARCH 1971

407
4. 11. 71
Lie

LIMBIC-DIENCEPHALIC MECHANISMS OF VOLUNTARY MOVEMENT[1]

C. H. VANDERWOLF[2]

University of Western Ontario

Data from a series of brain-behavior experiments utilizing techniques of ablation, intracranial stimulation, and electrical recording emphasize the importance of the concept of levels of function in understanding behavior. Initiation, performance, and cessation of voluntary behaviors (e.g., walking, manipulating objects) are associated with activation of ascending pathways from the diencephalon to the hippocampus and neocortex. More automatic behaviors (e.g., chewing, emotional expression) are not directly dependent on these pathways, even when they occur as conditioned responses. Electrical recordings indicate that hippocampal activity, in particular, is closely related to behavior. Wave morphology, amplitude, and frequency, are all related to different aspects of concurrent motor activity. Problems of the definition and mechanism of "voluntary" and "automatic" control of movement are discussed.

This review is an attempt to summarize several years of work on the behavioral functions of the medial thalamus and hippocampal formation and to speculate on the neural mechanisms of higher level control of movement.

In general, the results emphasize the importance of the concept of levels of function in understanding the neural control of behavior. Hughlings Jackson observed that voluntary movements are more affected by cerebral lesions than automatic movements (including reflexes and stereotyped or rhythmical activities), suggesting that the latter may normally be performed without the direct participation of higher centers. Contemporary research on the effects of forebrain damage in animals confirms Jackson's observations. Further, the correlation of behavior with the spontaneous electrical activity of the brain yields results consistent with Jackson's hypotheses. During voluntary types of movement such as locomotion or the manipulation of an object, there is an increased activation of the dorsal hippocampus and neocortex. This does not occur during behavioral immobility nor during the performance of automatic behaviors such as emotional expression or shivering. Therefore, there appear to be special cerebral mechanisms which are required for the performance of voluntary movements but are not necessary for the maintenance of a fixed posture or the performance of automatic movements.

There is no generally accepted objective definition of "volition" or of "voluntary

---

[1] Research reported here was supported by grants from the National Research Council of Canada. I am indebted to B. H. Bland, D. O. Hebb, D. Hopkins, Doreen Kimura, R. S. Sainsbury, and I. Q. Whishaw for helpful criticisms and suggestions.

[2] Requests for reprints should be sent to C. H. Vanderwolf, Departments of Psychology and Physiology, University of Western Ontario, London 72, Canada.

© 1971 by the American Psychological Association, Inc.

movement" and the difference between it and "automatic movement." Some possibilities are that voluntary acts are those that can be brought under the control of many different drive states and can be performed in a variety of combinations and sequences. Automatic motor activities comprise reflexes as well as a number of activities which are not obviously reflexive but occur only in association with particular drive states.

It is proposed that each distinguishable drive state is associated with the activation of a cluster of automatic responses unique to that drive and whose components are interconnected at a hypothalamic or lower brainstem level. Thus, "feeling cold" is associated with cutaneous vasoconstriction, shivering, a crouched posture, etc.; fear is associated with cutaneous vasconstriction, an exaggerated startle response, special patterns of urination and defecation, etc. Voluntary movements do not form an integral part of such response clusters, but can be temporarily coupled to any one of a large number of them. Thus, running may be controlled by hunger, thirst, fear, sexual or material drive, etc., on different occasions. Motor activities such as shivering or salivation are more rigidly connected with particular drive states.

The coupling between a drive state and a voluntary movement is assumed to occur via intermediary mechanisms in the forebrain which establish a "program" determining the selection of particular motor acts and the sequence in which they are to be performed. Overt behavior results when a program is activated by means of a trigger mechanism, largely located in the diencephalon. There may be an interaction between this movement triggering mechanism and a second system whose activation results in behavioral immobility. The occurrence of automatic motor activity is not directly dependent on these intermediary processes.

When different behaviors are deliberately conditioned in tests of learning, they appear to depend on the same response mechanisms as when they occur spontaneously. Improved understanding of the motor system may be fundamental to a correct interpretation of learning phenomena.

BEHAVIORAL EFFECTS OF MEDIAL THALAMIC DESTRUCTION

Behavioral studies have been carried out on a large number of rats in which the medial thalamic nuclei had been destroyed electrolytically. The lesions were relatively extensive, involving the medial and midline nuclei as well as the intralaminar nuclei and habenula. Details of most of the work are provided in earlier publications (Vanderwolf, 1962, 1963a, 1964, 1966, 1967b, 1968, 1969b).

The behavioral changes produced by large medial thalamic lesions can be grouped into three successive phases, a hyperactive phase, an akinetic phase, and a chronic "recovered" phase. Immediately following surgery, while the rats are still partially anaesthetized, they become extremely active, usually running forward or circling but occasionally jumping repetitively. A few animals eat or gnaw at protruding objects. These behaviors may occur almost continuously for several hours. One rat, placed in an activity wheel, ran approximately three miles in an hour and a half. Usually the rats were kept in plastic buckets or small cardboard boxes during the period of hyperactivity to prevent self-injury or fatal exhaustion. Seizures occurred in some cases. The day following surgery, all traces of hyperactivity had ceased, the rats having become almost totally motionless for long periods. Many rats did not eat or drink without assistance for 1–3 days, although reflexive components of feeding behavior were present. For example, if the wet spout of a water bottle was brought to the mouth, vigorous licking and biting would occur. If contact with the lips and tongue was broken, no compensating head movements were made to regain the spout and the licking and biting movements would stop, but could be restarted if the spout was presented a second time. In contrast, more voluntary acts such as spontaneous walking occurred rarely or not at all. Recovery from this akinetic phase was gradual, but within 2–3 weeks the rats were not easily distinguishable from sham operate control animals ("recovered" phase). They gained weight at a normal rate when food and water were continuously available, groomed themselves adequately, and did not exhibit any abnormality in motor performance. That is, such acts as walking, running, etc., were performed normally. At times, muscular tonus seemed somewhat subnormal throughout the entire body, but this was never sufficient to interfere with normal posture.

For several days following surgery, a few of the rats exhibited exaggerated defensive behavior somewhat like the syndrome produced by septal damage. Thus, a severely akinetic rat might scratch and bite savagely when handled. In order to minimze this type of behavior, rats were always housed in groups of 2–10 rats in a large cage. (Even normal rats become savage when kept in isolation for long periods—Hatch, Wiberg, Balazs, & Grice, 1963). The net result was that the medial-thalamus-damaged rats tended to be tamer than normal or sham operate rats which were similarly housed. This observation has been confirmed and extended by Wishart (1968) who used a rating scale technique to assess emotionality following medial thalamic lesions.

Behavioral testing was usually begun 2–3 weeks following surgery. An early finding was that medial-thalamus-damaged rats were almost totally unable to avoid an electric shock in a one-way avoidance situation. The task required that the rats learn to run from a black compartment, in which they had been manually placed, into a white one in less than 5 seconds. When first shocked in this test, the rats reacted in a relatively normal fashion, exhibiting startle, running, squealing, defecation, piloerection, exophthalmia, etc. However, after several training trials, the thalamus-damaged rats all became unusually docile and easy to handle as compared with normal rats undergoing the same treatment. When placed in the black (shock) compartment, they remained completely motionless until shock was presented, whereupon they ran rapidly off the grid. This behavior usually persisted unchanged even after several hundred training trials. That is, avoidance behavior was virtually absent, but escape behavior was relatively intact. This dissociation has now been demonstrated in a variety of avoidance test situations (Bohus & de Wied, 1967; Cardo, 1961; Delacour, Albe-Fessard, & Libouban, 1966; Olton & Isaacson, 1967; Thompson, 1963; Vanderwolf, 1962, 1963a). The effects are similar if medial thalamic lesions are made either before or after training or if the rats are tested in either a one-way or two-way apparatus (shuttle box). Further, the deficit does not disappear with time. Rats trained as long as a year after surgery show a typical failure to avoid.

Several factors should be considered in the interpretation of these results. Fibers of the spinothalamic system (pain afferents) terminate in the intralaminar nuclei (Bowsher, 1961; Magoun, 1963; Mitchell & Kaelber, 1967), and various types of fear or escape behavior can be elicited by stimulation in the medial thalamic area (Black & Vanderwolf, 1969; Olds & Olds, 1963; Roberts, 1962). Thus, the effect of medial thalamic lesions on avoidance behavior could be the result of reduced fear or an impairment of pain perception. It is also possible that impaired learning ability is responsible for the effect. Gastaut, Jus, Jus, Morrell, Storm van Leeuwen, Dongier, Naquet, Regis, Roger, Bekkering, Kamp, and Werre (1957) have suggested that the thalamic unspecific system (which was extensively damaged in the foregoing experiments) is the site of the "temporary connection" produced by conditioning procedures.

It is unlikely, however, for reasons given below, that any of these factors are responsible for the loss of active avoidance behavior observed in rats with medial thalamic damage. After several days of avoidance training, such rats often began to squeal when placed in the black (shock) compartment although they made no visible attempt to run away. They did not squeal (or squealed less) when placed in the white (safe) compartment. This indicates that some fear responses can be conditioned (in Pavlov's sense) in a rat with medial thalamic damage. A similar result was reported by Doty, Beck, and Kooi (1959) who found that in cats in an aversive conditioning situation, large medial diencephalic and midbrain lesions abolished a leg flexion response but not a respiratory conditioned response.

Autonomic responses to electric shock, such as defecation (Vanderwolf, 1962) and heart rate acceleration (Delacour & Santacana, 1967), can also be conditioned in rats following medial thalamic destruction. Further, medial thalamic lesions do not impair performance in tests of passive avoidance in which animals avoid an electric shock by remaining motionless (McNew & Thompson, 1966; Vanderwolf, 1962; Wishart, 1968). Finally, thalamic lesions placed in the posterior ventral nucleus raise the pain threshold, but do not necessarily impair avoidance performance. On the other hand, medial thalamic lesions can depress avoidance performance without raising the pain threshold for shock to the feet (Delacour & Santacana, 1967; Roberts & Carey, 1963).[3] Thus, the effect of medial thalamic damage on avoidance behavior is not the result of either defective pain perception or an inability to acquire conditioned fear responses.

Large medial thalamic lesions impair learning ability as measured by complex maze tasks (Delacour et al., 1966; Gross, Chorover, & Cohen, 1965; Warren & Akert, 1960). Therefore, it is conceivable that rats with large medial thalamic lesions perform poorly in avoidance tests requiring active movement because of inability to learn in this particular type of task, even though they can acquire vocal and autonomic conditioned responses or passive avoidance behavior. However, medial-thalamus-damaged rats give evidence of an ability to *learn* what to do in an aversive situation even though they fail to *perform* active avoidance responses. For example, large medial thalamic lesions do not prevent good performance in a three-choice brightness-discrimination task motivated by escape from electric shock (Vanderwolf, 1969b). Further, in the standard one-way test, medial-thalamus-damaged rats that fail to avoid, though usually totally motionless, sometimes make intention movements which appear to be the first stages of an avoidance response. For example, they may turn the head, or even the entire body (the rats must turn 180° before running), to face the safe compartment, even though running movements themselves do not appear. Similarly, if the task used requires jumping out of a box (11 inches deep) to avoid shock, many of the rats adopt a vertical posture in preparation for jumping, although the jump itself does not occur. If sufficient time is allowed in such cases, an avoidance response may eventually be made. Thus, avoidance performance can be improved considerably if the time allowed for a response is increased from 5 to 15 seconds (Vanderwolf, 1962).

It is perhaps worth emphasizing that rats with large medial thalamic lesions are not deficient in primary motor abilities. On those occasions when the rats actually make an avoidance response, the movements are normal. For example, in jumping out of a box, they leap accurately to the edge of the apparatus and climb out.

In conclusion, a rat with a large medial thalamic lesion is grossly deficient in the ability to avoid a painful shock even though it is not paralyzed in the usual sense and possesses adequate sensory and learning abilities and motivational capacity. The defect appears to lie in the initiation of movement, in dysfunction of a mechanism linking perceptual, mnemonic, and motivational systems to the voluntary motor system. There may be two distinct phases in higher level motor control. First, a programming mechanism must select from a large number of possible movements those which are appropriate in a given situation. The central representations of these movements must be maintained in a subthreshold state of excitation for some time in such a way that they can be activated in a particular sequential order (see Lashley, 1951; Milner, 1970, pp. 92–98). Then, at an appropriate moment, a trigger mechanism fires off whatever motor activities have been programmed. Since medial-thalamus-damaged rats usually perform appropriately in aversive situations, provided they move at all, the lesion appears to interfere more severely with the triggering than with selection of movement. Triggering can still be accomplished by an external stimulus, since the rat performs appropriately when shock is applied to the feet (escape behavior).

---

[3] The escape threshold for stimulation of the tooth pulp is raised by destruction of the parafascicular nucleus, a finding which may be related to the medial location of trigeminal afferents (Mitchell & Kaelber, 1967).

The existence of a movement-triggering mechanism involving the diencephalon is supported by experiments on intracerebral stimulation in conscious animals. There are points in the medial thalamus (Endröczi, Yang, Lissák, & Medgyesi, 1959) and posterior hypothalamus where electrical stimulation invariably produces motor activity of some sort, but the actual pattern may vary—running, jumping, climbing, digging, etc.—(from the same placement with the same stimulation parameters) depending on the environmental circumstances. In contrast, points in the midbrain tegmentum, for example, determine not only the occurrence but also the pattern of movement. Environmental factors have little effect. Thus, during forced circling elicited from the midbrain, a rat may be driven into obstacles or off the edge of a table (Vanderwolf, Hopkins, & Bland, unpublished observations). The behavior elicited in the first case corresponds to what one might expect from the activation of a movement trigger. In the latter case, a specific motor sequence appears to be activated more directly.

An important point in the interpretation of the effect of medial thalamic lesions is that not all motor patterns are equally affected. In an aversive situation, respiratory and autonomic responses can be conditioned readily in the experimental animals, but running and jumping cannot. Such phenomena call to mind the suggestion of Hughlings Jackson (see Taylor, 1958) that different motor activities can be arranged conceptually in a rough scale extending from highly automatic movements such as respiration, cardiac and vasomotor activity, and simple somatomotor reflexes, to highly voluntary behaviors among which were included speech, manipulation, and directed locomotion. Following Hughlings Jackson, one may say that medial thalamic lesions interfere more with voluntary behavior in a fear-producing situation than with the more automatic behaviors.

There is a difficulty. If medial thalamic lesions interfere with the operation of a mechanism which triggers voluntary movements, one would expect an effect on a variety of different behaviors. In fact, many apparently voluntary movements appear unaffected. Rats with large medial thalamic lesions perform well in simple appetitive tests and can easily press levers, run down alleys, etc., to obtain food or water (Replogle, 1960; Vanderwolf, 1962, 1967b). This behavior contrasts sharply with the poverty of spontaneous movement exhibited by the same animals in an aversive situation. However, several observations suggest that initiation of movement may be impaired by medial thalamic lesions even in nonaversive situations. When medial-thalamus-damaged rats, which had been resting quietly in their cages, were suddenly placed on a familiar tabletop, they did not begin to walk about as quickly as normal animals. They showed no sign of fear and became quite active within a minute or so. This effect was demonstrated quantitatively in two situations. In one test, lights and a buzzer were presented to animals in their home cages and the latency of an investigative response (rearing) was recorded. In a second test, the time taken to leave a small platform was recorded. Medial-thalamus-damaged rats were slow to initiate movement in both tests (Vanderwolf, 1969b).

A clue to the nature of the deficit produced by large medial thalamic lesions is provided by the immediate postoperative behavior of the experimental animals. The initial hyperactivity can be attributed to a discharging lesion (resulting from injury discharges, irritation by metallic deposits, etc.) in a movement-controlling system. The fact that behavioral seizures sometimes accompany the hyperactivity supports such a hypothesis. The ensuing akinesia may be attributable to a depression of the same system (in part only, since the rats are undoubtedly very fatigued) from which there is a gradual recovery during the next few weeks. Ultimately, only traces of the earlier akinesia can be observed, and the rats are actually hyperactive in some situations. Thus, medial thalamic lesions appear to produce a mild generalized impairment in the initiation of movement, but in a chronic preparation, the defect is obvious only in aversive situations.

Medial-thalamus-damaged rats which are actually hyperactive in some nonaversive situations may exhibit a total absence of spontaneous movement during avoidance training. Observation of normal rats helped to clarify this puzzling situation. When a rat is frightened, it may freeze, that is, remain alert but completely immobile for long periods. When head or locomotor movements occur, they often start or stop in an abrupt fashion, in contrast to the gradual buildup of such movements in an undisturbed rat. This suggests that during fear there is an alternation between two central states, one resulting in alert immobility, the other resulting in movement, perhaps by activation or deactivation of the trigger system discussed above. One might suppose, if the two states act reciprocally, that when activity in the trigger system is depressed by medial thalamic damage, the immobility-producing system would dominate behavior in aversive situations. The trigger system might still work fairly well in appetitive situations in which the immobility-producing system is not activated.

A simpler explanation than this might seem to be that the effect of medial thalamic damage is merely to increase the normal tendency to freeze (as a positive response) in emotional situations. Such an explanation is logically unsatisfactory, however, because the primary effect of a brain lesion cannot be to add something to the behavioral repertoire. It must always depress some function, even though other functions may be "released" and thus appear to be strengthened—a point made by Hughlings Jackson many years ago. In fact, in absolute terms, the tendency to freeze is probably reduced by a medial thalamic lesion. Rats with medial thalamic damage freeze less than normal during exploration of a novel environment (Vanderwolf, 1969b), and they also defecate less than normal under these circumstances (Vanderwolf, 1962). This, as well as the observation that a medial thalamic lesion produces a rat which is hyperactive in some situations and totally immobilized in others, suggests that more than one functional system is disrupted. One may speculate that medial thalamic damage raises the threshold of the central fear mechanism slightly in addition to its effect on the initiation of movement.

There is considerable evidence that the thalamic unspecific system contains at least two different functional subsystems, since it can exert either activating or deactivating influences on behavior and cortical electrical activity (Monnier, Hösli, & Krupp, 1963). For example, different workers have reported that avoidance responses can be either suppressed (Bloch & Hebb, 1956) or faciliated (Cardo, 1967) by stimulation of the medial thalamus. According to Buser (1966), the posterior intralaminar nuclei form part of an activating pathway ascending to the sensorimotor cortex, while the more rostral midline nuclei exert a deactivating influence. Both these sets of nuclei, as well as the thalamic dorsomedial nucleus, whose destruction is said to result in behavioral hyperactivity (Warren & Akert, 1960), were extensively damaged in the experiments discussed here. Therefore, complex effects on behavior are to be expected.

It has been shown that a variety of treatments including destruction of the subcallosal cortex or the septal nuclei (McCleary, 1961, 1966) and electroconvulsive shock (see Vanderwolf, 1963b) interfere with the ability of rats to remain motionless in such situations as tests of passive avoidance and conditioned suppression. This result has been taken to mean that the function of somatomotor inhibitory mechanisms was impaired. Presumably, the normal balance between the tendency to move and the tendency to remain motionless in an aversive situation was tipped in favor of movement. Perhaps the somatomotor inhibitory mechanisms of the forebrain exert an effect on behavior by modulating activity in the movement trigger system. If so, it might be possible to improve avoidance performance in rats with medial thalamic lesions by reducing activity in the movement-inhibiting system, thus tending to restore the normal balance between movement and immobility. This was attempted in two experiments. Medial-thalamus-damaged rats were tested on a one-way avoidance task, given electroconvulsive shock, septal lesions, or a control treatment, then retested on the one-way task. Both electroconvulsive shock and septal damage produced an improvement in performance in about half the rats tested, while control rats remained unchanged (Vanderwolf, 1964, 1968). These findings support the hypothesis that septal damage and electroconvulsive shock disrupt mechanisms which normally control activity in a movement trigger system.

The septal nuclei, like the medial thalamic nuclei, must form part of many functional systems. It is widely held that a major role of the septal nuclei is to inhibit somatomotor activity. However, Peacock and Hodes (1951) demonstrated a heavy concentration of points in the septal-preoptic area which facilitated movements produced by cortical stimulation. Other points in the same general region had an inhibitory effect (Hodes, Peacock, & Heath, 1951). It has also been shown that stimulation of the septal nuclei usually results in an increased discharge rate of cells in the centromedian nucleus of the thalamus (Sommer-Smith, Powarzynski, Stirner, & Grünberg, 1965). Thus, both somatomotor inhibitory and facilitatory systems appear to be present in the septal nuclei, and a large lesion might be expected to produce a complex behavioral change. This may account for the finding that although septal lesions usually have a "disinhibitory" effect on behavior in aversive situations (impairment of passive avoidance, facilitation of avoidance in a shuttle box, etc.), they produce an impairment in "one-way" tasks in which the rat must run or jump out of a box (Kenyon, 1962; Rich & Thompson, 1965), and they reduce spontaneous running in activity wheels (Dirlam, 1969; Douglas & Raphelson, 1966).

The hypotheses outlined here may be of some use in understanding a number of the phenomena associated with shock avoidance behavior in intact animals. For example, one might expect that engaging the animals in some sort of motor activity immediately before a test session might activate the trigger system enough to improve avoidance performance. In fact, pretest exploration of a novel environment does improve avoidance performance in both normal (Posluns & Vanderwolf, 1970) and medial-thalamus-damaged rats (Vanderwolf, 1966). Handling normal rats during the intertrial interval (producing struggling movements) has a similar effect (Wahlsten, Cole, Sharp, & Fantino 1968). The effect is probably not restricted to learned tasks. Activities such as exploration or swimming reduce response latency in a simple locomotor task (stepping from a small platform to a table) in which learning plays a minimal role (Vanderwolf, 1969b). However, the effect may be restricted to more voluntary motor patterns, since a conditioned vocal response was not facilitated by pretest exploratory behavior (Vanderwolf, 1966).

Normal rats trained in a two-way (shuttle box) avoidance test with the usual delay-conditioning procedure behave in many ways like medial-thalamus-damaged rats in a one-way test. Acquisition is improved by septal damage (King, 1958) or electroconvulsive shock (Vanderwolf, 1963b) or by increasing the time during which an avoidance response may occur (CS–UCS interval) (Black, 1963; Low & Low, 1962; Schwartz, 1958). The latter effect is attributable to better performance rather than more efficient learning (Anderson, 1969). One could speculate that the normal rats' poor performance in the two-way apparatus is due to an imbalance between immobility and movement triggering under these particular conditions. It is interesting that dogs, which show little tendency to freeze in fearful situations, perform much better than rats in a shuttle box, and do not improve their scores greatly when the CS–UCS interval is increased (Brush, Brush, & Solomon, 1955; Church, Brush, & Solomon, 1956). Both results suggest ready operation of the movement trigger. One might expect comparable results in other mammals which have little tendency to freeze, such as mustelids, for example. In contrast, one might expect that cats, which seem to have an efficient mechanism for remaining motionless when excited (e.g., when watching a mouse hole), would be poor performers in a shuttle box and would perform better as the CS–UCS interval was increased. In fact, cats do not perform well in a shuttle box, but it has not been shown that their performance can be improved by an increased CS–UCS interval, although the data reported by Schrier (1961) show a trend in that direction.

One of the main conclusions of the work on medial thalamic damage is that higher level motor control can be disrupted by a lesion which leaves perceptual-mnemonic processes relatively undisturbed. This point is demonstrated very clearly in certain neurological conditions in man. Lhermitte, Gauthier, Marteau, and Chain (1963) described a case in which an ischemic lesion had destroyed the core of the brainstem from the pontomesencephalic junction to the mammillary bodies and massa intermedia. The patient was almost totally mute and akinetic. She could, however, flex one arm and by this means could indicate the correct answers to rather complicated questions, indicating that perceptual and cognitive processes were relatively intact. A temporary but otherwise analogous condition occurs in sleep paralysis and cataplexy (Cave, 1931; Daniels, 1934; Wilson, 1928) and some types of petit mal attacks (Tizard & Margerison, 1963). During attacks, patients may be totally unable to move but are aware of ongoing environmental events as shown by a correct verbal report following recovery. Attacks of cataplexy are often precipitated by strong emotion (especially laughter or fright), a fact which suggests a parallel to the effect of medial thalamic damage in the rat.

A different type of akinesia with relatively good preservation of perceptual-mnemonic abilities is observed in severe Parkinson's disease. In this case, inability to initiate locomotion, for example, appears due to a disturbance of postural mechanisms. Patients may be able to stand, or even raise their feet one at a time, but fail to walk because they cannot lean forward or rock from side to side, as is required for normal locomotion. With assistance (being held forward slightly and rocked from side to side), they walk quite well (Martin, 1967). The deficit in such cases appears to lie more with the execution than with the triggering of movement, and appears to have little in common with the defect produced by medial thalamic damage.

The status of perceptual-mnemonic abilities in animals made totally akinetic by central lesions has never been adequately studied. Ingram, Barris, Fisher, and Ranson (1936) were of the opinion that cats with lesions of the posterior hypothalamic area were "conscious" although they did not move about. Such a hypothesis is consistent with the observation that cats made akinetic by posterior hypothalamic lesions may exhibit good activation of the cortical electroencephalogram (EEG) but no concomitant behavioral arousal (Feldman & Waller, 1962; Lindsley, Schreiner, Knowles, & Magoun, 1950).

RELATION OF DIENCEPHALIC AND HIPPOCAMPAL EEG TO SOMATOMOTOR ACTIVITY

Two general conclusions were suggested by the experiments on the behavioral effects of medial thalamic damage. First, that there are considerable differences in the neural organization of voluntary and automatic motor activities; second, that structures in the medial diencephalon form part of a trigger mechanism involved in the initiation of voluntary movement.

EEG recording has provided a supplement to ablation as a method of studying higher level control of movement. Clear relations were found between motor activity and the pattern of electrical activity in the dorsal hippocampus and medial thalamus (Vanderwolf, 1967a, 1969c; Vanderwolf & Heron, 1964). Most of the work was done on the dorsal hippocampus, and this will be discussed here.

It is useful to classify the EEG derived from the dorsal hippocampal pyramidal cells in the waking rat into three main types, as was done previously by Stumpf (1965b) in the rabbit. Rhythmical slow activity (RSA) consists of trains of nearly sinusoidal waves of a frequency varying from about 6 to 12 Hertz (Hz.). Amplitude varies widely depending on the exact location and orientation of the electrode tips with respect to the dipole of the pyramidal cell (Green, Maxwell, Schindler, & Stumpf, 1960), but under optimal conditions (when the tips of the electrode pair lie on either side of the cell layer), potentials of 2 millivolts or more can be obtained. At a given electrode site, both amplitude and frequency of RSA vary in relation to the behavior of the animal.

Large-amplitude irregular activity (LIA) is a second waveform which differs from RSA in that it has a component of very slow waves and lacks the rhythmical character of RSA. Large-amplitude spikes with a duration of 50–100 milliseconds may occur at irregular intervals.

A third waveform is small-amplitude irregular activity (SIA) which appears as a sudden suppression of hippocampal activity and ordinarily lasts only 1–2 seconds, in contrast to the two preceding types which may continue steadily for many minutes.

All three of the foregoing patterns will occur at a single placement in the hippocampus. The particular pattern exhibited depends on the behavior of the animal. Electrical activity in the dentate gyrus CA-4 area does not exhibit a comparable variation and, to visual inspection, consists of a continuous pattern of fast activity (mostly 15–50 Hz.) with an amplitude up to 1 millivolt. However, the dentate gyrus CA-4 area does exhibit an arousal reaction similar to the neocortex. Fast activity also appears in hippocampal (pyramidal cell layer) recordings, but often has a lower amplitude. At the same amplification, a hippocampal record contains more slow activity (less than 15 Hz.) than a dentate gyrus record, whether RSA or LIA is present.

If electrodes are placed at random in the hippocampal formation, mixtures of RSA (usually of rather small amplitude) plus a good deal of fast activity are very frequently encountered. Such patterns presumably represent an intermingling of activity from the hippocampus, dentate gyrus, and subiculum.

During drowsiness and slow-wave sleep, various other patterns of irregular large-amplitude slow waves, spikes, and spindles occur throughout the hippocampal formation. These waveforms will not be discussed in detail here.

The patterns of EEG activity in the dorsal hippocampus are related to concurrent motor activity in a consistent way. This correlation has been studied most extensively in the rat (Vanderwolf, 1969c; Whishaw & Vanderwolf, 1970) and guinea pig (Sainsbury, 1970), but has also been examined, in this laboratory, in the rabbit, Mongolian gerbil, and cat. Observations have been reported in the dog by Black, Young, and Batenchuk (1970), Dalton (1968), Lopes da Silva and Kamp (1969), Storm van Leeuwen, Kamp, Kok, and Tielen (1967), and Yoshii, Shimokochi, Miyamoto, and Ito (1966). Differences between the various species appear to lie mainly in the frequencies of EEG activity and in the ease with which good RSA can be obtained.

RSA accompanies voluntary types of movement, including walking, or running straight ahead, or backing up, turning, rearing, jumping, climbing, struggling when held, swimming (studied in rat only), head movements, postural changes (such as shifting weight from one foot to another, getting up, lying down, etc.), manipulation (such as handling a food pellet or pressing a lever in a Skinner box) and digging in sawdust. Within this group of motor activities, there is a relation between the extent of the movement (total motor units activated per unit time?) and the amplitude of the accompanying RSA. Rearing and jumping are associated with large-amplitude RSA, walking with somewhat smaller amplitude, handling an object with still smaller amplitude, and so on. Tiny movements of the head or one limb are associated with RSA of minimal amplitude, sometimes only 20%–25% of the maximum occurring at a particular placement. The amplitude of the RSA accompanying a given motor act is not rigidly fixed, but varies (as does the motor act itself) from one occasion to another. For example, a sudden fast run, as when a rat has been startled, is usually accompanied by a larger amplitude RSA than ordinary walking.

Variations in amplitude of RSA are probably responsible for the following phenomenon. At optimal electrode placements, acts such as walking or isolated movements of the head or limbs are always accompanied by RSA, but at mixed activity sites (RSA plus a great deal of fast activity) it may be found, for example, that while RSA is present on 100% of the occasions that a rat rears, it may appear on only 75% of the occasions that it walks, and almost never when it manipulates a piece of food or moves the head sideways slightly while standing still. In such cases, presumably, only large-amplitude RSA can be "seen" by the recording electrodes as a result of the location of the electrode tips with respect to the dipole of the pyramidal cells.

The frequency of hippocampal RSA usually varies over a range of about 6–12 Hz. in the rat. Somewhat surprisingly, there is very little relation between the amplitude and period of individual waves, even when samples are taken during a wide range of behaviors, suggesting that wave period (frequency) and amplitude are controlled by different mechanisms. The results discussed above suggest that the amplitude of the RSA is related to the gross amount of concurrent motor activity. Frequency increases, on the other hand, are associated with the initiation of movement. The latter point can be clearly demonstrated in a shock-avoidance task which requires rats to jump out of a box. This task has several advantages: (*a*) the moment of initiation of muscular activity can be defined accurately, (*b*) a large movement is required, ensuring large-amplitude RSA, and (*c*) the use of shock reinforcement suppresses irrelevant spontaneous motor activity, permitting the observation of a sudden transition from immobility to vigorous movement. Using well-trained rats which exhibit a minimum of irrelevant motor activity immediately preceding the jump response, it can be shown that RSA with a frequency of 6–7 Hz. precedes the jump by several seconds. Between 1 and 2 seconds before the jump, the frequency rises, reaching a peak value, usually 8–10 Hz., just before or at the instant that overt motor activity begins. Similar frequency shifts accompany the initiation of running, whether it occurs as an avoidance response, or spontaneously, or in a motor-driven activity wheel. If motor activity continues steadily once it has been initiated, RSA frequency declines, although the mean frequency remains somewhat higher than during the period of preparation for movement. This has been demonstrated in a motor-driven activity wheel where a rat can be made to run as long as 8 hours without stopping. Aside from frequency shifts associated with the onset of running, RSA persists unchanged throughout this period (Whishaw & Vanderwolf, 1970). Thus, initiation of movement is accompanied by a high-frequency burst of RSA, but RSA continues at a lower frequency as long as movement continues.

Finally, it has been found that small movements, such as handling food or moving the head slightly without other movement, are accompanied by RSA of a lower mean frequency than is seen during major movements, such as walking or rearing. The relation of this to the other findings has not yet been determined.

There is no evidence that the RSA accompanying a given movement is influenced by the motivational or stimulus factors which may have given rise to that movement. For example, the RSA accompanying walking seems much the same during exploration, food getting, shock avoidance, or forced running in a motor-driven activity wheel.

Behavioral immobility is usually accompanied by LIA. A rat standing motionless on four legs, on two legs, or hanging motionless by the forepaws from a support will exhibit this EEG pattern. Thus, maintenance of a static condition of muscular tonus is not associated with RSA.

A number of motor patterns are associated with hippocampal LIA which is similar, perhaps identical, to the LIA occurring during behavioral immobility. These motor patterns are commonly regarded as relatively automatic and include licking, chewing, chattering the teeth (as a component of defensive behavior in rat and guinea pig), champing without anything in the mouth, copious salivation (observed in one rat which got a bit of dry food stuck in its throat), piloerection (observed in rat and guinea pig), urination, defecation, pelvic thrusting and ejaculation during coitus (studied in guinea pig only), face washing, licking and biting the fur on any part of the body, scratching with one hind foot, vocalization (squealing in rat, barking in dog, etc.), shivering (observed in rat and guinea pig), and stamping on the ground with the hind feet (observed in the gerbil by Kramis & Routtenberg, 1969). As in the case of RSA, the reason a given act is performed has no obvious influence on the pattern of LIA exhibited in the hippocampus. A similar waveform appears during licking, regardless of whether the thing licked is water, the animals' own fur, or the vulva of a female in estrus.

It often happens that a motor pattern which (when performed alone) is accompanied by RSA is performed simultaneously with a motor pattern which (when performed alone) is accompanied by LIA. In such cases, RSA always occurs. Examples are numerous. Movements of the vibrissae, as isolated acts, are accompanied by LIA, but if they occur together with head movements or walking (the usual situation) the combination is accompanied by RSA.[4] A male guinea pig while copulating exhibits LIA during pelvic thrusting, but if obliged to step forward with the hind limbs or use the forelimbs to restrain the movements of the female, RSA appears. When a rat scratches its flank, it stands motionless except for the movements of the active leg. Continuous LIA appears in the hippocampus. However, when scratching the head, the rat assumes a rather unstable posture in which the neck and trunk are curved laterally. Small postural adjustments occur continually, and the hippocampal record contains considerable small-amplitude RSA.

Yawning appears to be a rather automatic act in man, and therefore one might expect it to be accompanied by hippocampal LIA in the rat (see also p. 98). However, yawning is associated with stepping movements or walking in rats and the total behavioral pattern is associated with RSA. Whether yawning, as an isolated act, is associated with RSA is difficult to determine. The righting reflex is another example of this type. Although the act of righting undoubtedly has a reflex basis, if an animal is placed on its back, RSA is always present as it struggles and as it rolls over. One may speculate that in an intact animal, voluntary movements collaborate with a brainstem reflex in the restoration of normal posture.

Another factor which can pose difficulties in interpretation is that RSA amplitude is small when very slight voluntary movements are performed. Therefore, it might be argued that licking, for example, is really accompanied by RSA, but of such small amplitude that it cannot be identified easily. In the case of blinking, for example, this argument may be valid, but for the most part it has been clear that tiny head or paw movements (e.g., manipulation) only 1–2 millimeters in extent are associated with clear RSA, while much larger movements such as biting, face washing, shivering, and pelvic thrusting are accompanied by LIA. Thus, there appear to be two distinct classes of behavior of which one is accompanied by hippocampal RSA while the other is not.

SIA occurred infrequently under the conditions studied so far and its behavioral correlates are less well known. It appears to occur on occasions when an ongoing movement is abruptly stopped. It has been observed under the following circumstances: when drowsy rats were startled so that they leapt to their feet but did not run; in rats in an avoidance test, jumping to the edge of a box and almost going over the other side (requiring a sudden halt of forward movement); in rats in a Skinner box, thrusting their heads into the food cup and stopping abruptly to seize a food pellet which had just been delivered; in rats in a motor-driven activity wheel where behavior consisted alternately of jumping (with RSA) and clinging to the wheel (with SIA). A simple demonstration can be made in the gerbil. If a gerbil is running, a whistle will sometimes make it stop suddenly. The stop is accompanied by a transition from RSA to SIA, then back to RSA as the gerbil begins to move again. If the gerbil remains motionless for some time, SIA is replaced by LIA. It appears that SIA is associated only with an abrupt halt of ongoing movement and that when an animal stops moving ordinarily, SIA does not appear.

---

[4] According to Komisaruk (1970), vibrissae movements are correlated with RSA and tend to occur in phase with it. However, there are instances when RSA is present while the vibrissae are motionless, as in swimming, and other instances when the vibrissae move and RSA is absent.

BRAINSTEM CONTROL OF HIPPOCAMPAL ACTIVITY AND OF MOVEMENT

The findings described in the preceding section indicate that three types of hippocampal rhythm can be identified by simple inspection in an EEG record. One type is LIA which is present during alert immobility and during the performance of automatic movements. A second type is RSA which precedes and accompanies the performance of voluntary types of movement. Large-amplitude RSA accompanies extensive movements; smaller amplitude RSA accompanies lesser movements. Extensive movements occur in association with rises in RSA frequency. The third type of hippocampal activity, SIA, is associated with a sudden halt of ongoing movement, suggesting a relation to somatomotor inhibitory mechanisms.

Many of these findings can be related to existing anatomical and physiological knowledge. Green and Arduini (1954) demonstrated that electrical stimulation of the brainstem reticular formation, medial thalamus, or hypothalamus produced RSA in the hippocampus. Septal lesions abolished the effect, indicating an ascending pathway through the septal nuclei. Later physiological studies (Corrazza & Parmeggiani, 1963; Petsche, Stumpf, & Gogolák, 1962; Torii, 1961; Yokota & Fujimori, 1964) have indicated that the ascending pathway runs through the medial hypothalamus and medial preoptic area to the medial septal nucleus and, eventually, to the hippocampus. An additional pathway probably runs forward through the medial thalamus (Eidelberg, White, & Brazier, 1959; Kawamura, Nakamura, & Tokizane, 1961). Anatomical evidence confirms the existence of such ascending pathways, although details remain rather obscure (see reviews by Adey, 1959; Green, 1964; Lewis & Shute, 1967, Stumpf, 1965a).

It has been found that stimulation at some subcortical sites produces a suppression or flattening of hippocampal activity, a reaction which resembles the SIA pattern observed during normal behavior, although the electrically induced pattern can be made to continue for a longer period. This type of effect has been observed following high-frequency stimulation of the septum (Brücke, Petsche, Pillat, & Deisenhammer, 1959), lateral hypothalamus, and lateral preoptic area (Torii, 1961), possibly suggesting the existence of a second, antagonistic, ascending functional system running parallel to the RSA pathway from the brainstem to the hippocampus. This system might be responsible for normal SIA. Activity in both systems can be controlled from midbrain as well as diencephalic levels. Stimulation of the dorsolateral tegmentum yields RSA in the hippocampus; stimulation of the ventromedial tegmentum leads to a flattening of the hippocampal record (Torii, 1961). These effects probably result from an ascending activation of the diencephalic mechanism, since large hypothalamic lesions abolish RSA completely when the midbrain is intact; elimination of the midbrain does not abolish RSA (Kawamura et al., 1961; Kawamura & Domino, 1968).

In addition to RSA or hippocampal flattening, increased fast activity (15–50 Hz.) may appear in the hippocampus during brainstem stimulation. According to Stumpf (1965a), fast activity occurs independently of other hippocampal patterns of activity and can still be produced by brainstem stimulation after septal destruction, suggesting a totally separate functional system.

Activation of the RSA or hippocampal flattening (SIA?) systems has opposite effects on somatomotor activity. Stimulation of points in the amygdala, preoptic area, and ventromedial tegmentum which suppresses hippocampal slow activity with or without increased fast activity, depresses monosynaptic and polysynaptic spinal reflexes. Points in the medial hypothalamus, medial preoptic area, and dorsolateral tegmentum produce hippocampal RSA when stimulated and facilitate spinal reflexes (Yokota & Fujimori, 1964). Analogous effects can be demonstrated in a conscious animal. Ito (1966) found that septal stimulation produced both behavioral arrest and a flattening of hippocampal activity. Bland and Vanderwolf (1970a) have shown that unilateral stimulation of the dentate gyrus CA-4 area suppresses hippocampal slow activity bilaterally and produces an immediate arrest of ongoing locomotion, bar pressing, etc., but does not affect respiration, heart rate, shivering, or lapping water. No seizure discharges were detected, and the rats moved about normally when stimulation was stopped. Stimulation of the hippocampus at the same voltage does not have a comparable effect, although arrest can be obtained at higher voltages, suggesting a spread of current to the dentrate gyrus. Therefore, the hippocampus and dentate gyrus may have different functional roles in the control of voluntary movement.

Stimulation of the paramedian hypothalamus produces both hippocampal RSA and motor activity of a nonforced type as discussed above. That is, the movements produced may be running forward, turning, jumping, etc., depending on the environmental circumstances. Hypothalamic points that fail to produce hippocampal RSA upon stimulation also fail to produce significant voluntary motor activity (Bland & Vanderwolf, 1970b).

Brainstem mechanisms control both the frequency and amplitude of hippocampal RSA, although it has not been demonstrated that different control systems are involved, as the behavioral work suggests. However, Whishaw (1970) has recently found that the two parameters can be manipulated independently. Cooling a rat reduces RSA frequency without affecting amplitude. During the period of postictal depression following hippocampal seizure activity, RSA frequency is normal but amplitude is much reduced.

The frequency and amplitude of RSA elicited by brainstem stimulation vary, depending on the site stimulated. Kawamura and Domino (1968) have shown that RSA produced by medial thalamic stimulation has a lower frequency and smaller amplitude than RSA elicited by hypothalamic stimulation. The frequency of RSA is also related to the intensity of brainstem stimulation. Stronger stimuli yield higher frequency RSA at the onset of stimulation than weaker stimuli (Sailer & Stumpf, 1957; Yoshii et al., 1966).

If the posterior hypothalamus is stimulated in a freely moving animal, the onset of running behavior is accompanied by a high-frequency burst of RSA very similar to that accompanying the normal initiation of movement (Bland & Vanderwolf, 1970b). Thus, both the initiation of movement and the increase in RSA frequency may be due to activity in an ascending brainstem system. This conclusion is consistent with the hypothesis of a diencephalic movement trigger mechanism discussed above. Further, continued performance of a movement pattern is associated with lower frequency RSA than initiation of the same movement. This suggests that triggering of a movement is accomplished by a burst of activity in the diencephalic system which then continues to discharge at a reduced rate as movement continues. The experiments on medial thalamic lesions suggested a similar conclusion. During avoidance testing, it was often observed that when a rat had once begun to run, the movements could be completed without difficulty.

Medial thalamic lesions appear to interfere with the initiation of voluntary types of movement, but the effect is relatively slight. Posterior hypothalamic-subthalamic area lesions have a more severe effect, amounting to chronic akinesia (Magoun, 1950). The more ventral lesions also prevent the appearance of RSA in the hippocampus, while medial thalamic lesions may result only in a slight slowing of the rhythm (Kawamura et al., 1961). More rostrally, this relation disappears. Septal lesions, which prevent the appearance of RSA in the hippocampus, or removal of the entire hippocampal formation, do not result in akinesia. The defect produced by such lesions consists, in part, in a failure to organize motor activity effectively rather than an inability to initiate a motor act (see below pp. 105–106). In terms of the concepts presented here, one might say that the hippocampal formation takes part in the programming of an intended act but that no actual behavior occurs until the program is activated by the diencephalic trigger.

According to the views presented here, the presence of RSA in the hippocampus indicates an activated state. In opposition to this, it has often been suggested that the hippocampus has a predominantly inhibitory role with respect to behavior (Douglas, 1967; Kimble, 1969; McCleary, 1966) and that the presence of RSA indicates that the hippocampus is inactive, that is, that a structure which inhibits the brainstem is itself inhibited (Grastyán, Lissák, Madarász, & Donhoffer, 1959). This view is based partly on the finding that animals with large lesions of the hippocampal formation are "disinhibited" (i.e., hyperactive, performing poorly in tests of passive avoidance, tending to continue performing inappropriate movements, etc.) and partly on the assumption that regular "synchronous" activity necessarily indicates an inactive state. The latter assumption may be unjustified. The mammalian olfactory bulb, for example, generates very regular waveforms when an air blast or strong odor is presented, suggesting that strong excitation may favor the development of synchronous activity in a neurone population (Adrian, 1942, 1950). Many hippocampal units fire in phase with the RSA waveform (Green, 1964). As RSA frequency increases during the initiation of movement, such units presumably fire more rapidly, indicating an activation of function of some sort. Finally, RSA can be recorded from the thalamus and hypothalamus as well as the hippocampus. It seems improbable that all these structures would be inhibited during active movement.

The fact that "disinhibition" of behavior occurs when the hippocampal formation is damaged does not prove that the hippocampus normally plays an inhibitory role. The behavioral change may be partly due to removal of the dentate gyrus, since both structures are usually destroyed in a surgical lesion. Electrical stimulation of the dentate gyrus CA-4 area suppresses voluntary movement; stimulation of the hippocampus has no obvious effect unless seizure activity occurs. These facts could mean that it is the dentate gyrus rather than the hippocampus which has an inhibitory role. The finding that the hippocampal pyramids generate RSA while the dentate gyrus CA-4 area apparently generates fast activity supports the hypothesis of a functional differentiation of the two (also see Jackson, 1968). However, it should be borne in mind that the methods used in behavioral studies to date are too gross to permit demonstration of possible functional differences between the dentate gyrus, CA-4, or the various fiber systems that traverse this area.

There may be considerable functional differentiation in the septal nuclei as well as in the hippocampal formation. Donovick (1968) found that only those septal lesions which abolished RSA in the hippocampus produced impairment of discrimination reversal learning. Lesions anywhere in the septal area facilitated shuttle-box performance, regardless of whether RSA was abolished or not. This may suggest that the ascending RSA pathway plays a role in the organization of complex behaviors, such as discrimination reversal learning, but is not concerned with the inhibitory influences believed to be responsible for poor shuttle-box performance. Further, lesions of the medial septal area (presumably blocking hippocampal RSA) may interfere with the production of motor activity, since they depress wheel running and make rats "placid" (Clody & Carlton, 1969) as well as interfere with avoidance performance in a test in which the rat must jump out of a box (Rich & Thompson, 1965). Perhaps it is the more lateral septal structures which are concerned with inhibitory effects and whose destruction results in ferocity, impaired passive avoidance behavior, etc.

RELATION OF HIPPOCAMPAL RSA TO AUTONOMIC ACTIVITY

There is evidence that hippocampal RSA is associated with cardiovascular changes as well as with voluntary types of movement. Torii and Kawamura (1960) observed that spontaneous occurrences of RSA in a curarized cat were associated with sharp rises in blood pressure. Hypothalamic stimulation at points which produced hippocampal RSA also produced rises in blood pressure. However, stimulation at points in the amygdala, septal nuclei, and rostral hypothalamus resulted in a flattening of the hippocampal record and a lowering of blood pressure. At a number of points in the brainstem, Yokota and Fujimori (1964) observed a relation between pressor effects, hippocampal RSA, and facilitation of spinal reflexes on the one hand, and depressor effects, flattening of the hippocampal record, and inhibition of spinal reflexes on the other hand. In an aversive-conditioning situation, Black and de Toledo (1971) often observed a relation between the appearance of RSA and increased heart rate in curarized dogs.

An anatomical basis for correlations between hippocampal RSA and cardiovascular activity is suggested by evidence (Ban, 1966) that a sympathetic zone occupies a paramedian position throughout the hypothalamus and preoptic area, apparently overlapping extensively with the ascending RSA pathway. A parallel parasympathetic zone occupies a more lateral position in regions where electrical stimulation produces a flattening reaction in the hippocampus as well as depression of spinal reflexes.

These findings are consistent with the view that there is a close association of somatomotor and cardiovascular control in the central nervous system (Rushmer & Smith, 1959). Stimulation of the motor cortex produces not only movements, but also a rise in systemic blood pressure and cardioacceleration, as well as a shift in distribution of blood from the viscera to the somatic musculature. Renal blood flow, for example, is reduced. A similar constellation of effects can be obtained from points in the amygdala, hypothalamus, and midbrain (Bard, 1960; Fulton, 1949; Gellhorn, 1967; Hoff, Kell, & Carrol, 1963; Smith, Rushmer, & Lasher, 1960). The circulatory phenomena are due, at least in part, to a direct central nervous influence on the heart and vessels rather than to a secondary effect of muscular activity, since they can be demonstrated in curarized animals. Thus, there appears to be a synkinesis between somatomotor and vasomotor control analogous to the relation between convergence of the eyes, miosis, and contraction of the ciliary muscle. During movement, or preparation for movement, the central nervous system apparently activates the vasomotor system as well as the respiratory and general somatic musculature, as was suggested by Krogh and Lindhard (1913, 1917) many years ago. Once movements have been initiated, of course, many other factors such as postural changes, muscle pumping, input from sensory receptors in the heart, vessels, joints, and muscles, as well as metabolic changes, will also affect the circulation.

An interesting aspect of the relation between hippocampal RSA and behavior is that the central rhythm may precede movement by several seconds. As movement becomes imminent, RSA frequency rises. Similar anticipatory changes occur in the respiratory and circulatory systems, as was pointed out by Krogh and Lindhard and many subsequent investigators (Mountcastle, 1968, pp. 542–543). The heart rate tends to rise in anticipation of spontaneous walking in dogs or in anticipation of sucking movements in human infants (Gottlieb & Simner, 1966; Perez-Cruet & Gantt, 1959).[5]

A relation between movement and sympathetic activity is also indicated by the dilatation of the pupil which occurs in association with key pressing, speech, lifting weights, etc. (Hakerem & Sutton, 1966; Nunnally, Knott, Duchnowski, & Parker, 1967; Simpson & Paivio, 1966, 1968). Slight mydriasis can precede actual movement and can also occur when no visible movement is performed, as when a subject simply imagines something. Even in this case, however, minimal activity may be generated in movement-producing systems. According to numerous investigations, imagining a visual scene or imagining the performance of some motor act is accompanied by tiny eye movements, muscular activity in the limbs, etc. (Whatmore & Kohli, 1968).

---

[5] Simner (1966) reported that heart rate decreased prior to vocalization in the chick. Since chicks tend to stand still while making distress calls, it seems possible that the fall in heart rate was associated with immobility rather than the onset of vocalization. Unfortunately, movement was not recorded.

Thus, although there is evidence of increased sympathetic activity when a motionless animal develops RSA and begins to move, in general there is probably no simple relation between RSA and sympathetic activity. The cardiovascular system has evolved primarily to meet the metabolic requirements of active tissue, and during vigorous motor activity, the heart rate, for example, is high regardless of whether RSA is present or not. Thus, heart rate is higher during face washing than during ordinary walking in the rat, even though face washing is not accompanied by RSA (Vanderwolf & Vanderwart, 1970).

It is widely accepted that the cardiovascular system, in common with other autonomically innervated structures, is powerfully affected by emotional states, apparently without relation to overt movement. However, it seems possible that many "emotional" effects on the autonomic system are, in fact, related to the performance of special patterns of movement, but that the relation has gone unrecognized because autonomic activity is often preparatory, occurring well in advance of actual movement. For example, it seems obvious that penile erection is part of a general preparation for copulatory behavior, and that salivation can be a preparation for feeding. The cardiovascular activity characteristic of fear resembles the pattern seen during severe exercise, suggesting a preparation for rapid or prolonged running behavior. Urination and defecation during fear can be thought of in much the same way. Perhaps these activities have evolved in relation to fear because they are a means of reducing the abdominal contents, permitting freer movement of the diaphragm during maximal motor activity.

AROUSAL, ATTENTION, AND SLEEP

There is evidence that in the neocortex, electrical activity is related to behavior in somewhat the same way as it is in the hippocampus. Clemente, Sterman, and Wyrwicka (1964) and Roth, Sterman, and Clemente (1967) have reported that slow waves and spindles appear in the neocortex of a cat drinking milk or grooming, but do not occur during walking. In this laboratory, these waveforms have been observed (in the rat and guinea pig) in the hippocampal formation as well as the neocortex during chewing, face washing, shivering, or immobility, but have never been observed during the performance of behavior associated with hippocampal RSA. Thus, these observations are consistent with the general rule that hippocampal RSA is associated with low-voltage fast activity in the neocortex, while irregular hippocampal activity *may* be accompanied by slow waves and spindling (Green & Arduini, 1954).

It may seem contradictory that slow activity appears in the neocortex following the initiation of milk drinking in cats but does not appear when the cats eat solid food (Grandstaff, 1969). The difference between the two behaviors is probably that head movements (associated with low-voltage fast cortical activity and hippocampal RSA) occur almost continuously during eating in cats but occur only occasionally during milk drinking. Rats, unlike cats, chew their food, and apart from the continuous jaw movements, are often quite motionless while eating. Therefore, during chewing, spindling and slow waves often appear in the neocortex.

Another line of evidence is that the neocortical surface positive dc potential, measured against a reference electrode in the frontal bone in the rat, is large during behavioral immobility or face washing but smaller during walking, rearing, or head movement (Caspers, 1961). With the use of a transcortical recording technique, phasic surface positive "activation waves" have been reported in the neocortex of the cat during head movements but not during scratching, licking, or yawning (Calvet, Calvet, & Langlois, 1965).

Further, when a man clenches his fist, there is a blocking of the 25-Hz. beta rhythm which occurs in the precentral gyrus during immobility. If the fist remains clenched for more than about 1 second, the beta rhythm returns, only to disappear again when the grip is relaxed (Jasper & Penfield, 1949). Thus, the motor cortex in man, like the hippocampus in lower animals, seems to be activated during voluntary movement but not during the maintenance of a fixed posture.

All these observations indicate that during voluntary movement there is more activity in the ascending nonspecific projections to the neocortex, as well as to the hippocampus, than is present during behavioral immobility or automatic types of movement. A possible interpretation of this is that the neocortex and hippocampus play an active role in voluntary behavior but are less crucially involved in automatic behavior or in the maintenance of an unchanging posture. Such an interpretation assumes that neocortical slow waves and spindling and hippocampal LIA or spindling are indicative of an inactive functional state.

Other lines of evidence also suggest that normal activity in the hippocampus or neocortex is less important for automatic than for voluntary behaviors. Local seizure activity produced by stimulation of the hippocampus of a cat, while the animal was drinking milk, did not interrupt the ongoing lapping movements (Akert, 1961). Hippocampal seizure discharges, however, do prevent the occurrence of a more voluntary act such as an instrumental avoidance response (Flynn & Wasman, 1960; MacLean, Flanigan, Flynn, Kim, & Stevens, 1955). Dentate CA-4 stimulation arrests voluntary movement but not shivering or lapping water (Bland & Vanderwolf, 1970a). The different susceptibility of automatic and voluntary motor activities to diencephalic damage has already been discussed (see also Dirlam, 1969; Glickman, Higgins, & Isaacson, 1970).

The findings discussed here are relevant to the classical concepts of the relation of "arousal" to behavior (Hebb, 1955; Lindsley, 1951; Malmo, 1959). It might be suggested that behaviors such as chewing, face washing, and licking the back occur during states of low arousal or relaxation and are associated with hippocampal and cortical deactivation for this reason. However, it is important to distinguish cerebral activation from behavior arousal (Feldman & Waller, 1962; Lindsley et al., 1950). Face washing and licking of the back, in particular, are vigorous muscular activities in a rat and can be shown to have a higher heart rate than walking or moving the head, behaviors which are always associated with neocortical and hippocampal activation (Vanderwolf & Vanderwart, 1970). Also, as the work discussed here shows, if a frightened rat or guinea pig freezes or chatters its teeth, or shivers in response to lowered body temperature, the cortex and hippocampus show signs of deactivation (slow waves, spindling, and LIA) even though the animal would ordinarily be regarded as aroused at such times.

Hippocampal RSA is probably not associated with changes in attention in any way. Slight environmental stimuli will make a rat pause momentarily during the performance of such activities as lapping water or washing the face. Presumably this indicates "attention," yet the ongoing pattern of hippocampal LIA usually shows no alteration whatever. Of course, RSA will appear if the stimulus results in a head movement, postural adjustment, locomotion, etc. Similar results appear in a conditioning situation. RSA does not occur in dogs during the process of discriminating one conditioned stimulus from another, unless the animals move at the same time (Dalton & Black, 1968; Young & Black, 1970).

It would be interesting to know whether perceptual processes might not generally be more efficient when cortical mechanisms are not occupied with the control of voluntary movement. In man, it is well known that close attention is correlated with behavioral immobility, and according to J. O'Keefe (personal communication, 1969), cells in the cuneate nuclei are actually less excitable during the performance of behaviors associated with hippocampal RSA than during behaviors associated with LIA.

Hippocampal RSA appears during sleep (activated or paradoxical sleep) as well as in the waking state. An interpretation of this is suggested by evidence that there is a great deal of activity in the motor system during activated sleep but that the spinal motoneurone and its associated reflex circuitry are "switched off" by descending inhibitory activity, making gross movement impossible. The descending motor barrages result only in tiny twitches of the ears, vibrissae, paws, tail, etc. (Pompeiano, 1967). Whishaw and Vanderwolf (1970) found that during activated sleep, hippocampal RSA has a higher frequency and amplitude during periods of "twitching" than in the intervening quiet intervals. This suggests that motor activity is triggered by diencephalic-limbic mechanisms during activated sleep in much the same way as in the waking state.

In a number of situations, low-frequency RSA can be observed in the absence of movement in waking animals. As a rat prepares to jump out of a box, RSA is present, but the jump occurs only when the frequency of the RSA reaches 8 Hz. or more. In rabbits, sensory stimuli can produce long trains of RSA without any accompanying visible movement. Such RSA is always of lower frequency than is seen during actual voluntary movement (Harper, 1968), suggesting that, here too, a movement (running?) was "programmed" but that the frequency necessary for its initiation was not achieved. Cats also may exhibit low-frequency RSA at a time when they are not obviously moving about (Brown, 1968).

Animal hypnosis is yet another state in which low-frequency RSA is associated with absence of movement, but in addition there is a loss of the normal righting reflex. There may be an inhibition of spinal motoneurones in this case, as in paradoxical sleep (Harper, 1968; Klemm, 1966, 1969; McBride & Klemm, 1969).

HIPPOCAMPAL ACTIVITY AND LEARNING

It has been reported that the hippocampal EEG is altered during the course of training in a learning task. According to Grastyán et al. (1959), the initial response to a novel stimulus is a desynchronization of hippocampal activity. As conditioning proceeds, RSA appears in association with the prominent orienting movements ("vicarious trial and error" in American psychological writings) which occur early in acquisition (Tolman, 1948). After additional training, as the learned response becomes well established, RSA ceases to appear. These findings might appear contradictory to other reports that hippocampal activity consists largely of RSA during first exposure to novel stimuli (Radulovački & Adey, 1965) and that RSA does not disappear even after prolonged overtraining (Adey, 1962, 1967; Adey, Dunlop, & Hendrix, 1960). Nonetheless, Adey and his colleagues agree that changes in hippocampal activity do occur as a result of training, since a well-established approach response is accompanied by RSA of a higher frequency than is seen in the same animal early in training (Elazar & Adey, 1967).

However, other evidence suggests that the hippocampus may play a critical role in motor performance but is not essential to learning in lower animals. Flynn and Wasman (1960) showed that if conditioning trials are presented during hippocampal seizure activity, performance (leg flexion) is blocked, but learning is still possible apparently, since the animals perform quite well when subsequently tested in the normal state.

The changes in hippocampal electrical activity during training can also be accounted for on the hypothesis that such activity is related to movement rather than the learning process itself. For example, whether an animal exhibits hippocampal RSA or not when a novel stimulus is presented depends on the type of behavior produced. Exploratory behavior is associated with RSA; behavioral immobility usually is not. Early in training, in the type of apparatus (alley) used by Adey, an animal is likely to spend part of its time standing motionless but moving its head about. Later it is more likely to walk promptly toward the food reward. Since head movements in an otherwise motionless animal are associated with lower frequency RSA than walking, regardless of whether the movements are deliberately "conditioned" or not, it may be that the frequency shift during training is due to a change in behavior per se which is not necessarily related to learning.

It is also clear that mere training or repetition of a task does not lead to disappearance of RSA if it was initially present. Whishaw and Vanderwolf (1970) have recorded hippocampal activity in rats forced to run in a large motor-driven wheel for periods of up to 8 hours per day for several days in succession. RSA remained virtually unchanged throughout this period. Further, RSA appears during feeding behavior when a rat manipulates a large piece of solid food, a behavior which the animal performs hundreds of times every day of its adult life.

However, under other circumstances, RSA may be reduced or even seem to disappear during the course of a learning experiment. It was pointed out above that extensive voluntary movements are associated with higher amplitude RSA than smaller movements. If the recording electrodes are not placed in an optimal position and orientation with respect to the dipole of the pyramidal cells, RSA may be present only occasionally during the performance of extensive movements and seldom or never during the performance of small movements. At such electrode placements, RSA may "disappear" during training as the extensive irrelevant movements characteristic of the early phases of acquisition drop out. Such "disappearance" was demonstrated by Vanderwolf and Whishaw (1969) who showed that RSA was often present at "poor" hippocampal placements (little RSA present) on the first day of training in a bar-press task. On subsequent days, RSA was much less common as the rats began pressing regularly and stopped walking about. At "good" hippocampal placements, RSA did not disappear, but was reduced in amplitude as the rats' behavior changed from walking and rearing to regular bar pressing.

These results indicate that hippocampal RSA is related to the higher level control of movement and is not a sign of the activity of the learning process. A final line of evidence which supports this statement is that automatic behaviors can occur as conditioned responses in the absence of RSA. Posluns (1962) found that rats drugged with chlorpromazine do not make voluntary shock-avoidance responses although conditioned vocal responses may still occur. Hippocampal recordings taken from rats given chlorpromazine during avoidance testing (Vanderwolf, 1969a) showed that when the rats vocalized, no RSA appeared. However, RSA was still present if the rats walked or jumped, regardless of whether the behavior was a conditioned response, an escape response, or a spontaneous movement. Young and Black (1970) report that LIA occurs in the hippocampus of rats performing a licking response to avoid shock.

Thus, an automatic conditioned response does not appear to be controlled by the same neural mechanisms as a voluntary conditioned response. EEG findings confirm the results of the studies on the effect of medial thalamic lesions. In fact, there is considerable evidence that the performance of automatic behaviors as conditioned responses is less dependent on forebrain structures than more voluntary learned responses. In addition to points already mentioned, hippocampal seizure activity may have a more severe effect on a conditioned leg flexion response than on a conditioned respiratory response (MacLean et al., 1955–1956). This is also true of seizures initiated by application of acetylcholine to the neocortex (Fleming, 1963). Similarly, conditioned cardiac responses are still possible following removal of the neocortex (Bloch-Rojas, Toro, & Pinto-Hamuy, 1964) or cortical spreading depression (Morgenson & Peterson, 1966), although a behavioral avoidance response is impaired or abolished.

DEFINITION OF VOLUNTARY AND AUTOMATIC MOVEMENT

Various theoretical schemes have recognized differences in the level of organization of different behaviors. A distinction between reflexive and voluntary behavior has long been widely accepted (Fearing, 1930). Sherrington (1961, pp. 322–330) distinguished precurrent from consummatory acts and emphasized the dependence of the former on the distance receptors and the dependence of the latter on contact receptors. An analogous distinction was made between appetitive and consummatory behavior by Craig (1966). It is often emphasized that appetitive behavior initiates a behavioral sequence, while consummatory behavior terminates it.

The writings of Hughlings Jackson provide the most extensive existing treatment of levels of function of the nervous system (see Taylor, 1958). Jackson proposed that simple reflexes and stereotyped rhythmical movements, such as breathing, were controlled by the lower levels of the nervous system, while the more varied voluntary movements were dependent on the operation of higher, more recently evolved structures, especially the cortex. Lower (more automatic) motor centers could be controlled (either excited or inhibited) by higher centers, which could be controlled by still higher centers, and so on, the whole forming a multilevel hierarchical system. An important feature of Jackson's ideas was that a given function might be represented many times at different levels. Thus, one might say that the movement of flexing a limb is represented at the lowest level as a spinal reflex. At a higher level, limb flexion is integrated with other movements to permit walking, as may be observed in a decerebrate animal. At a still higher level, limb flexion or walking is integrated with other sensorimotor activities to form an adaptive behavior such as food getting. If a higher center is destroyed, or otherwise rendered nonfunctional, the next highest center is "released" and behavior is correspondingly simplified and made more automatic.

For reasons of historical priority, as well as easy relation to present day neurological and neurophysiological concepts, it has seemed advisable to adopt Jackson's terms of "voluntary" and "automatic" to refer to motor organization of different degrees of complexity. In opposition to this, it might be argued, for example, that behaviors associated with hippocampal RSA should be called "appetitive" while those associated with hippocampal LIA are "consummatory." However, this type of distinction is often difficult to apply and is not easily related to questions of physiological mechanism. Threat behaviors such as piloerection and chattering of the teeth in a guinea pig are accompanied by LIA but might be regarded as "appetitive" since they precede fighting. Therefore, Hughlings Jackson's terminology seems preferable.

Nonetheless, although everyone understands the meaning of the term "voluntary" in an intuitive way, it is difficult to define objectively. It seems probable that there is no single identifying characteristic. In common usage, voluntary acts are those controlled directly by an intention or idea, whereas a reflexive act is one which is elicited and controlled by some definite stimulus (see Hebb, 1966). However, such activities as salivation or penile erection can be aroused solely by imagination, but are nonetheless not fully voluntary since they cannot be performed readily on verbal command and become totally unavailable during fear.

One factor suggested by the preceding examples may be that motor patterns differ in the degree to which their occurrence is dependent on the presence of a particular motive state. When someone "feels cold," shivering movements occur and, when severe, cannot be suppressed (Denny-Brown, Gaylor, & Uprus, 1935; Uprus, Gaylor, & Carmichael, 1935). Shivering does not occur during the state of "feeling warm" and cannot then be made to occur by voluntary effort. The fact that salivary secretion is depressed in man during strong fear has been put to practical use in several ancient methods of lie detection (Smith, 1967). Yet, on occasions related to eating, salivation occurs and cannot be suppressed. A large class of somatic and autonomic activities, including emotional expression, yawning, vomiting, components of feeding, eliminative, and sexual behavior, etc., resemble these examples in that they share one or both of the following characteristics—they are difficult to suppress when the appropriate motive state is present and difficult or impossible to produce when some other motive state is present. In contrast, a voluntary type of movement can be easily brought under the control of any one of a number of different motive states. For example, in man, walking or manipulative behavior can be brought into the service of food getting, attack, flight, reproduction, etc., but laughing and weeping are not equivalently available to all motive states. Everyone knows that during weeping it is all but impossible to prevent sobbing, keep the corners of the mouth up, etc., and also that during a gay mood, these motor patterns cannot be reproduced convincingly. Laughter is much the same. Giggling can be almost uncontrollable, especially in young people. On the other hand, attempts to feign such behaviors when one is not amused are described as a "strained smile," "forced laugh," etc.

Such phenomena suggest a dual-level control of facial expression in man. There appear to be a number of species-specific patterns of emotional expression which develop without specific training or imitation (Darwin, 1965, Goodenough, 1932; Thompson, 1941), although the *occasions* on which a given pattern occurs are obviously very much influenced by learning (Ekman, Sorenson, & Friesen, 1969). With increasing age and experience (Fulcher, 1942), the facial musculature comes under the control of an additional system which permits voluntary movements, including acting and the feigning or concealment of emotions which is so characteristic of man, but also occurs in chimpanzees (Hebb & Thompson, 1954). The two systems can be dissociated by brain damage. Patients in whom the cortical motor pathway is believed to be bilaterally damaged may laugh or cry uncontrollably, with abundant facial movement, yet be incapable of showing the teeth upon command. A hemiplegic patient may retain full emotional expression on both sides of the face although voluntary facial movements are largely restricted to one side. In contrast, following brainstem damage, voluntary facial movement may be preserved while emotional expression is lost (Brown, 1967; Ironside, 1956; Wilson, 1924). Thus, there appears to be a subcortical mechanism for "genuine" emotional expression, but when facial movements are to be controlled by some other motive, such as the desire to comply with a verbal request, a cortical mechanism is brought into play.

This type of dual control is not restricted to "emotional expression" as conventionally understood. Shivering can be mimicked to some extent by a normal person. In cases of hemiplegia, such voluntary control is presumably lost unilaterally, although the shivering in response to cold is normal on both sides of the body (Uprus et al., 1935).

It seems possible that a great many motor patterns in nonhuman animals are of the type which is restricted to one or a small number of motive states. Breland and Breland (1961, 1966) report a large number of observations on the difficulty of training a hungry pig not to root or a chicken not to scratch prior to the delivery of food in operant training situations. Such a requirement may be analogous to asking a hungry man not to salivate when food is presented. Further, Breland and Breland did not succeed in teaching a goat to bleat or a cow to run in order to obtain food, perhaps because these behaviors are not a part of normal feeding behavior in these animals.

Kellogg (1968) reports that it is very difficult to train a chimpanzee to vocalize to obtain food even though the animal has a wide repertoire of "spontaneous" calls (mostly linked to various emotional states). Nonetheless, the chimpanzee is very teachable in other respects, and Gardner and Gardner (1969) have successfully taught one to communicate using more than 30 signs of American sign language (the language used by the deaf in North America).

Bolles (1969) found that rats do not easily learn to rear to avoid shock, although running is rapidly acquired. It contrast, it is well known that rats can be quickly taught to rear to obtain food or water, even in classroom demonstrations. Grooming behavior in rats, unlike head movements or locomotion, is not susceptible to punishment or escape training in which noise is used as a reinforcing stimulus (Bolles & Seelbach, 1964).

As another type of example, it appears to be common knowledge among farmers and ranchers that horses in winter will dig through snow with their forefeet to reach the buried grass. Cattle do not appear to do this and do not fare as well unless fed by man. Nonetheless, cattle can easily acquire simple instrumental responses (such as pumping water for themselves by pressing a panel with their muzzles) and are quite capable of making the movement of pawing the ground, since bulls do this as part o their threat behavior. Apparently, the movements of pawing cannot be easily connecte with the central representation of hunger i cattle.

In conclusion, it seems likely that many motor patterns in animals resemble shivering, emotional expression, penile erection, etc., in man in that certain patterns will be performed whenever a given motive state is present in sufficient intensity, but not at other times. The actual structure of motor control in different species is almost totally unknown. Breland and Breland (1966, p. 87) suggest that higher animals (porpoises as well as primates) are more susceptible to "cross-drive conditioning" than lower forms.

Many of the phenomena of drive-linked behavior might be explained by assuming that different motive states each possess a unique cluster or constellation of relatively automatic somatic and autonomic effector activity, adapted to the particular behavior which is to be performed. A given component of such a reaction cluster may be quite rigidly linked to the total cluster, as shown by consistent appearance when the drive state is sufficiently intense and failure to appear in association with other drive states. Such complex reaction patterns are sometimes regarded as reflexes. In Pavlovian physiology, the readiness of an animal to take food, to salivate, etc., is regarded as an indication of the excitability of a feeding reflex or feeding center. The term "center" in this instance does not refer to a particular locus but to a multicomponent system organized at all levels of the neuraxis (Bykov, 1958, p. 525). Elaborations of an essentially similar idea are found in ethological theory and have proved extremely useful in the analysis of instinctive behavior (Craig, 1966; Tinbergen, 1951). The concept of activation of an innate response system can also be applied to many conditioned behaviors. It has long been known that when a neutral stimulus is repeatedly paired with electric shock, for example, the "conditioned response" includes a vast constellation of visceral and somatic phenomena which resemble the response to the shock itself. However, voluntary types of learned behavior cannot be interpreted in this way. An innate feeding or defensive reflex can hardly include all the motor patterns (walking, jumping, manipulation, etc.) which animals will perform to obtain food or avoid electric shock. Perhaps, when an animal learns to perform an act such as running to obtain food, the central representation of the running pattern becomes coupled to the central representation of feeding in some way and is controlled by it, so that temporarily the running response behaves in much the same way as an innately coupled motor pattern such as salivation. The comparisons presented by Kimble (1961) show that running or bar pressing for food follow much the same rules of acquisition, generalization, extinction spontaneous recovery, etc., as a salivary conditioned response. If the feeding system (including salivation, chewing, etc.) is activated artificially by hypothalamic stimulation in a satiated animal, a previously learned voluntary motor pattern will also appear (Maire, 1965; Miller, 1958). Such facts suggest that the two types of response share the same sensory and motivational control mechanisms. The control mechanisms cannot be identical however, since automatic behaviors, occurring as conditioned responses, are often less affected by forebrain damage than voluntary behaviors (see above, p. 101) and may occur independently of the RSA system. According to the point of view advanced here, the difference lies in the greater complexity of the motor organization of voluntary behaviors, one aspect of which is dependence on forebrain control systems.

Some of the sources of this complexity have already been suggested. Some sort of switching circuitry must be necessary to control voluntary movement patterns if they are to be connected with different drives on different occasions. The more automatic components of the cluster of motor activities associated with a drive could be interconnected more simply, since little flexibility is required. Complex circuitry must also be required to achieve the varying sequential pattern of voluntary movement. In man, a given movement of the hand can be preceded or followed by any of a very large number of movements, but it is difficult to produce any rearrangement whatever of some automatic sequences of muscular contractions, such as swallowing. The programming circuitry which permits voluntary movements to be brought under the control of a drive and performed in an effective sequence is presumably located largely in the cerebral hemispheres, in the neocortex, hippocampal formation, etc. Presumably, a program does not result in actual behavior unless the diencephalic movement trigger is also activated, the effect of the two summating to produce overt motor activity.

According to this interpretation, the presence of RSA in the hippocampus during walking, for example, may indicate the activation of a cerebral mechanism which temporarily couples the central representation of walking to the particular motive system dominant at the moment and permits regulation of rate, direction, etc. An instant later, running or head movement might be so coupled. Analysis of stimulus input and activation of memory engrams might be in progress as well, but these processes do not seem to require activation of the RSA system. During an automatic act such as shivering, RSA does not occur (one may speculate) because shivering is innately coupled to a cold defense motive system at a brainstem level, and the cerebral mechanism becomes superfluous.

However, not all movement patterns which occur in the absence of RSA are restricted to a single motive state. Licking, for example occurs during feeding, grooming, sexual, and maternal behavior. Rats can learn to lick a water spout to avoid shock (Young & Black, 1970). Biting and gnawing seem much the same. Despite this apparent flexibility, cerebral mechanisms are not activated during continuous licking or biting, since LIA occurs in the hippocampus and slow waves and spindling may occur in the neocortex. In explanation of this, it is possible that both licking and biting comprise a number of distinct motor patterns, each of which is innately coupled to a particular motive state. Biting the fur during grooming, for example, occurs at a much faster rhythm than biting and chewing food. Another possibility is that acts like licking and biting are not under the direct control of the cerebrum at all, but are operated indirectly, as reflexes.

An example of the indirect control of a reflex by a higher level mechanism is swallowing in man. Movements of the tongue (relatively voluntary) press food, saliva, etc., into the pharyngeal afferent zone of the swallowing reflex. Swallowing results. If the mouth is completely dry and empty, or if the pharyngeal surface is painted with cocaine, afferent stimulation is reduced and swallowing becomes difficult or impossible (Pommerenke, 1928). This type of control of reflex activity may occur fairly widely. McDonnell and Flynn (1966) have shown that the "killing bite" produced by hypothalamic stimulation in the cat is abolished if afferents from the lips are sectioned. Further, semicomatose rats with medial thalamic lesions (pp. 84–85) often lick and bite vigorously at a water spout applied to the lips at a time when they do not move the head in pursuit of the spout, even when it is moved only a few millimeters away. Chronic mesencephalic rats exhibit similar reflex licking and biting (Woods, 1964). Thus, when a normal rat licks or bites some object, it may do so by voluntary positioning of the body and head in such a way that the lower level automatisms are activated reflexively. If so, training a rat to lick water to avoid shock involves training voluntary movements of the head and limbs. Licking occurs automatically when the water spout is contacted.

The interpretations offered here may help account for the different sensitivity of voluntary and automatic behavior patterns to many types of forebrain damage. An automatic motor pattern may be coordinated largely by brainstem mechanisms in a normal animal. Presumably, it can be activated relatively directly by sensory input, and therefore is minimally dependent on the cerebral hemispheres. In a decerebrate rat, walking and head movements as well as licking and biting occur automatically (Woods, 1964). However, a decerebrate rat does not walk toward a food pellet and bite off a piece, presumably because the mechanisms that bring walking and head movement under the joint control of hunger and sensory input are absent. More restricted lesions may sometimes exert their effects by disrupting the control drive states normally exert on voluntary movement. For example, Kimble, Rogers, and Hendrickson (1967) found that following hippocampal lesions, parturient female rats did not gather their pups into the nest and hover over them as normal females do. The rats still possessed all the components of maternal behavior, but the total pattern was disordered. One rat, for example, would execute normal hovering behavior some distance removed from her pups. It is conceivable that such behavior occurs because the maternal drive is no longer able to control activity in the locomotor system and so bring the rat to its nest and pups when the hovering response has a low threshold.

OPERANT AND CLASSICAL CONDITIONING

A common distinction between voluntary and involuntary motor activities is based on the differences between smooth and striped muscles and their respective innervations. Since 1800 or even earlier (Fulton, 1949), what is now called the autonomic system has been associated with involuntary vegetative activities, while the neural systems controlling striped muscles have been associated with voluntary movement. A modern version of this idea is the suggestion that autonomic activities are susceptible to classical (Pavlovian) conditioning, while somatic activities are conditionable by instrumental or operant procedures.

Such ideas are not supported by present day knowledge. Neurological as well as everyday experience indicates that some somatic motor activities (e.g., shivering, sneezing, etc.) are not easily subjected to voluntary control. On the other hand, various autonomic activities such as piloerection or heart rate can sometimes be controlled in a voluntary manner in man or by operant training in animals (Kimmell, 1967; Lindsley & Sassaman, 1938; Miller, 1969).

It seems probable that operant control of autonomic activity can often be understood in the context of control of movement in general. It is generally impossible to elicit either somatic or autonomic reflexes in isolation (Fulton, 1949, p. vii), and cardiovascular activity in particular is closely linked to somatic motor activity (see above). In conditioning situations, increases in heart rate are associated with movement, decreases in heart rate with immobility (Obrist, 1968; Obrist & Webb, 1967; Obrist, Webb, & Sutterer, 1969; Roberts, 1970) in both appetitive and aversive situations. Behavioral experiments, apparently, have not established that emotion, motivation, attention, etc., have any effect on heart rate independent of their effects on movement.

It is sometimes assumed that the effect of movement can be eliminated in experiments on conditioning heart rate, blood pressure, etc., simply by curarizing the experimental subjects. However, as pointed out above, the motor system can probably alter heart rate or blood pressure directly without prior muscular activity. In fact, the occurrence of hippocampal RSA in a curarized animal is accompanied by increased heart rate and other signs of sympathetic activity. Now, operant control of RSA can be achieved in curarized dogs, and if the animals are later retested in the normal state, it is found that they have learned to struggle, toss the head, etc., in response to the CS (Black et al., 1970; Dalton, 1968). Heart rate, blood pressure, and renal blood flow can all be altered by operant procedures (Miller, 1969), and since all these variables can be controlled by stimulation of forebrain motor systems, it seems possible that the conditioning phenomena are dependent on the curarized animal learning to struggle, etc. However, it should be remembered that increases in heart rate, blood pressure, or respiration can precede actual movement, as Krogh and Lindhard pointed out many years ago. It might be possible to obtain operant control of such anticipatory motor processes and so obtain increased heart rate, for example, in the absence of any obvious somatic motor activity. However, the relation to motor activity might still be detectable by hippocampal recording.

Traditional distinctions between operant and classical conditioning now seem difficult to maintain. According to Hebb (1956), even the classical salivary conditioned response should be regarded as an operant, since an animal must salivate in order to eat. Similarly, the leg flexion response in classical aversive conditioning may be a type of voluntary trial-and-error behavior.

It is sometimes suggested that operant and classical conditioning should be distinguished on the basis of the procedures used. This type of distinction seems irrelevant. A valid classification should be based on the biological properties of the responses considered. Further, since the distinction is purely logical, there is no reason why it should not be continued indefinitely, dividing and subdividing learned behavior into many types, all based on minor procedural differences in training.

On the other hand, consideration of the response required, and a distinction between higher and lower levels of control, may often aid in understanding what actually happens in a conditioning situation. For example, during any conditioning procedure, some autonomic motor activities may appear as part of the response cluster associated with a drive. Others may occur in preparation for, and as an accompaniment of, voluntary somatomotor activity. Thus, a rat trained to run an alley for food might show increased heart rate in anticipation of running (voluntary), and salivation in anticipation of eating (drive response cluster). If the animal had learned to sit motionless to get food, the heart rate would presumably be lower but salivation would still occur. During eating, or subsequent face washing, heart rate would also be elevated, but this presumably would be due to lower level automatic regulation, since the RSA system would not be active.

During learning, voluntary motor activity may become coupled to an active drive state and may also be brought under the control of specified sensory stimuli. How all this is accomplished is almost a total mystery. The suggestion of Glickman and Schiff (1967) that activation of species-specific motor patterns is crucial to reinforcements seems readily compatible with the proposals made here. A discussion of possible neurological mechanisms of reinforcement is provided by Milner (1970).

This discussion has emphasized that while different drives may control voluntary movement via a hippocampal mechanism, hippocampal RSA is not, of itself, a correlate of motivational or reinforcement processes. Evidence supporting the latter conclusion includes the following: (*a*) RSA is not present during reinforcing activities such as chewing, licking, etc., (*b*) animals trained to remain motionless by an avoidance procedure or by the use of intracranial stimulation do not exhibit RSA (Dalton & Black, 1968; Paxinos & Bindra, 1970), and (*c*) brainstem points which produce hippocampal RSA when stimulated can be neutral, rewarding, or punishing in tests of self-stimulation (Ito, 1966; Kramis & Routtenberg, 1969; Routtenberg & Kramis, 1968). In all these situations RSA remains correlated with voluntary movement, however.

REFERENCES

ADEY, W. R. Recent studies of the rhinencephalon in relation to temporal lobe epilepsy and behavior disorders. *International Review of Neurobiology*, 1959, **1**, 1–46.

ADEY, W. R. EEG studies of the hippocampal system in the learning process. In P. Passouant (Ed.), *Physiologie de l'hippocampe*. Paris: Éditions du Centre National de la Recherche Scientifique, 1962.

ADEY, W. R. Hippocampal states and functional relations with cortico-subcortical systems in attention and learning. In W. R. Adey & T. Tokizane (Eds.), *Progress in brain research: Structure and function of the limbic system*. Amsterdam: Elsevier, 1967.

ADEY, W. R., DUNLOP, C. W., & HENDRIX, C. E. Hippocampal slow waves: Distribution and phase relationships in the course of approach learning. *American Medical Association Archives of Neurology*, 1960, **3**, 74–90.

ADRIAN, E. D. Olfactory reactions in the brain of the hedgehog. *Journal of Physiology*, 1942, **100**, 459–473.

ADRIAN, E. D. The electrical activity of the mammalian olfactory bulb. *Electroencephalography and Clinical Neurophysiology*, 1950, 2, 377–388.

AKERT, K. Diencephalon. In D. E. Sheer (Ed.), *Electrical stimulation of the brain*. Austin: University of Texas Press, 1961.

ANDERSON, N. H. Variation of CS–US interval in long-term avoidance conditioning in the rat with wheel turn and with shuttle tasks. *Journal of Comparative and Physiological Psychology*, 1969, **68**, 100–106.

BAN, T. The septo-preoptico-hypothalamic system and its autonomic function. In T. Tokizane & J. P. Schadé (Eds.), *Progress in brain research. Correlative neurosciences. Part A: Fundamental mechanisms*. Amsterdam: Elsevier, 1966.

Bard, P. Anatomical organization of the central nervous system in relation to control of the heart and blood vessels. *Physiological Reviews,* 1960, **40**, Suppl. No. 4, 3–26.

Black, A. H. The effects of CS–US interval on avoidance conditioning in the rat. *Canadian Journal of Psychology,* 1963, **17**, 174–182.

Black, A. H., & de Toledo, L. The relationship among classically conditioned responses: Heart rate and skeletal behavior. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II; Current research and theory.* New York: Appleton-Century-Crofts, **1971**, in press.

Black, A. H., Young, G. A., & Batenchuk, C. Avoidance training of hippocampal theta waves in Flaxedilized dogs and its relation to skeletal movement. *Journal of Comparative and Physiological Psychology,* 1970, **70**, 15–24.

Black, S. L., & Vanderwolf, C. H. Thumping behavior in the rabbit. *Physiology and Behavior,* 1969, **4**, 445–449.

Bland, B. H., & Vanderwolf, C. H. Effects of dentate gyrus stimulation on behavior and hippocampal electrical activity. *Proceedings of the Canadian Federation of Biological Societies,* 1970, **13**, 47. (a)

Bland, B. H., & Vanderwolf, C. H. Effects of hypothalamic stimulation on behavior and hippocampal EEG. Paper presented at the meeting of the Eastern Psychological Association, Atlantic City, N. J., April 1970. (b)

Bloch, V., & Hebb, D. O. Étude des phénomènes d'enrayment et d'activation du comportement par stimulation thalamique et reticulare chez le rat non-anaesthésié. *Psychologie Française,* 1956, **1**, 8–9.

Bloch-Rojas, S., Toro, A., & Pinto-Hamuy, T. Cardiac versus somatomotor conditioned responses in neodecorticate rats. *Journal of Comparative and Physiological Psychology,* 1964, **58**, 233–236.

Bohus, B., & de Wied, D. Avoidance and escape behavior following medial thalamic lesions in rats. *Journal of Comparative and Physiological Psychology,* 1967, **64**, 26–29.

Bolles, R. C. Avoidance and escape learning: Simultaneous acquisition of different responses. *Journal of Comparative and Physiological Psychology,* 1969, **68**, 355–358.

Bolles, R. C., & Seelbach, S. E. Punishing and reinforcing effects of noise onset and termination for different responses. *Journal of Comparative and Physiological Psychology,* 1964, **58**, 127–131.

Bowsher, D. The termination of secondary somatosensory neurons within the thalamus of Macaca Mulatta: An experimental degeneration study. *Journal of Comparative Neurology,* 1961, **117**, 213–222.

Breland, K., & Breland, M. The misbehavior of organisms. *American Psychologist,* 1961, **16**, 681–684.

Breland, K., & Breland, M. *Animal behavior.* New York: Macmillan, 1966.

Brown, B. B. Frequency and phase of hippocampal theta activity in the spontaneously behaving cat. *Electroencephalography and Clinical Neurophysiology,* 1968, **24**, 53–62.

Brown, J. W. Physiology and phylogenesis of emotional expression. *Brain Research,* 1967, **5**, 1–14.

Brücke, F., Petsche, H., Pillat, B., & Deisenhammer, E. Die Beeinflussung der "Hippocampus-arousal-Reaktion" beim Kaninchen durch elektrische Reizung im Septum. *Pflügers Archiv für die Gesamte Physiologie,* 1959, **269**, 319–338.

Brush, F. R., Brush, E. S., & Solomon, R. L. Traumatic avoidance learning: The effects of CS–US interval with a delayed-conditioning procedure. *Journal of Comparative and Physiological Psychology,* 1955, **48**, 285–293.

Buser, P. Subcortical controls of pyramidal activity. In D. P. Purpura & M. D. Yahr (Eds.), *The thalamus.* New York: Columbia University Press, 1966.

Bykov, K. M. (Ed.) *Textbook of physiology.* Moscow: Foreign Languages Publishing House, 1958.

Calvet, J., Calvet, M. C., & Langlois, J. M. Diffuse cortical activation waves during so-called desynchronized EEG patterns. *Journal of Neurophysiology,* 1965, **28**, 893–907.

Cardo, B. Rapports entre le niveau de vigilance et le conditionnement chez l'animal. *Journal de Physiologie,* 1961, **53**, 1–212.

Cardo, B. Effects de la stimulation du noyau parafasciculaire thalamique sur l'acquisition d'un conditionnement d'évitement chez le rat. *Physiology and Behavior,* 1967, **2**, 245–248.

Caspers, H. Changes of cortical D.C. potentials in the sleep-wakefulness cycle. In G. E. W. Wolstenholme & M. O'Connor (Eds.), *The nature of sleep.* London: Churchill, 1961.

Cave, H. A. Narcolepsy. *American Medical Association Archives of Neurology and Psychiatry,* 1931, **26**, 50–101.

Church, R. M., Brush, F. R., & Solomon, R. L. Traumatic avoidance learning: The effects of CS–US interval with a delayed-conditioning procedure in a free-responding situation. *Journal of Comparative and Physiological Psychology,* 1956, **49**, 301–308.

Clemente, C. D., Sterman, M. B., & Wyrwicka, W. Post-reinforcement EEG synchronization during alimentary behavior. *Electroencephalography and Clinical Neurophysiology,* 1964, **16**, 355–365.

Clody, D. E., & Carlton, P. L. Behavioral effects of lesions of the medial septum of rats. *Journal of Comparative and Physiological Psychology,* 1969, **67**, 344–351.

Corazza, R., & Parmeggiani, P. L. Central course of the afferent system eliciting the theta rhythm upon sciatic stimulation. *Helvetica*

Physiologica et Pharmacologica Acta, 1963, **21**, C10–C12.

Craig, W. Appetites and aversions as constituents of instincts. In T. Verhave (Ed.), *The experimental analysis of behavior.* New York: Appleton-Century-Crofts, 1966.

Dalton, A. Hippocampal electrical activity in operant conditioning in the dog. Unpublished doctoral dissertation, McMaster University, Hamilton, Ontario, 1968.

Dalton, A., & Black, A. H. Hippocampal electrical activity during the operant conditioning of movement and refraining from movement. *Communications in Behavioral Biology*, 1968, **2**, 267–273.

Daniels, L. E. Narcolepsy. *Medicine*, 1934, **13**, 1–122.

Darwin, C. *The expression of the emotions in man and animals.* Chicago: University of Chicago Press, 1965. (Orginally published: London: Murray, 1872.)

Delacour, J., Albe-Fessard, D., & Libouban, S. Rôle chez le rat de deux noyaux thalamiques dans le conditionnement instrumental. *Neuropsychologia*, 1966, **4**, 101–112.

Delacour, J., & Santacana, M. P. Rôle du thalamus médian dans l'établissement et le retention de conditionnements défensifs classiques. et instrumentaux. *Neuropsychologia*, 1967, **5**, 237–252.

Denny-Brown, D., Gaylor, J. B., & Uprus, V. Note on the nature of the motor discharge in shivering. *Brain*, 1935, **58**, 233–237.

Dirlam, D. K. The effects of septal, thalamic, and tegmental lesions on general activity in the hooded rat. *Canadian Journal of Psychology*, 1969, **23**, 303–314.

Donovick, P. J. Effect of localized septal lesions on hippocampal EEG activity and behavior in rats. *Journal of Comparative and Physiological Psychology*, 1968, **66**, 569–578.

Doty, R. W., Beck, E. C., & Kooi, K. A. Effect of brain-stem lesions on conditioned responses of cats. *Experimental Neurology*, 1959, **1**, 360–385.

Douglas, R. J. The hippocampus and behavior. *Psychological Bulletin*, 1967, **67**, 416–442.

Douglas, R. J., & Raphelson, A. C. Septal lesions and activity. *Journal of Comparative and Physiological Psychology*, 1966, **62**, 465–467.

Eidelberg, E., White, J. C., & Brazier, M. A. B. The hippocampal arousal pattern in rabbits. *Experimental Neurology*, 1959, **1**, 483–490.

Ekman, P., Sorenson, E. R., & Friesen, W. V. Pan-cultural elements in facial displays of emotion. *Science*, 1969, **164**, 86–88.

Elazar, Z., & Adey, W. R. Spectral analysis of low frequency components in the electrical activity of the hippocampus during learning. *Electroencephalography and Clinical Neurophysiology*, 1967, **23**, 225–240.

Endröczi, E., Yang, T. L., Lissák, K., & Medgyesi, P. The effect of stimulation of the brain stem on conditioned reflex activity and on behavior. *Acta Physiologica Academiae Scientarum Hungaricae*, 1959, **16**, 291–297.

Fearing, F. *Reflex action: A study in the history of physiological psychology.* Baltimore: Williams & Wilkins, 1930.

Feldman, S. M., & Waller, H. J. Dissociation of electrocortical activation and behavioural arousal. *Nature*, 1962, **196**, 1320–1322.

Fleming, D. E. Response arrest associated with the corticifugal projection of chemically induced local seizure discharge. *Electroencephalography and Clinical Neurophysiology*, 1963, **15**, 84–90.

Flynn, J. P., & Wasman, M. Learning and cortically evoked movement during propagated hippocampal after discharges. *Science*, 1960, **131**, 1607–1608.

Fulcher, J. S. "Voluntary" facial expression in blind and seeing children. *Archives of Psychology*, 1942, **38**, 1–49.

Fulton, J. F. *Physiology of the nervous system.* New York: Oxford University Press, 1949.

Gastaut, H., Jus, A., Jus, C., Morrell, F., Storm van Leeuwen, W., Dongier, S., Naquet, R., Regis, H., Roger, A., Bekkering, D., Kamp, A., & Werre, J. Étude topographique des réactions électroencéphalographiques conditionées chez l'homme. *Electroencephalography and Clinical Neurophysiology*, 1957, **9**, 1–34.

Gardner, R. A., & Gardner, B. T. Teaching sign language to a chimpanzee. *Science*, 1969, **165**, 664–672.

Gellhorn, E. *Principles of autonomic-somatic integrations.* Minneapolis: University of Minnesota Press, 1967.

Glickman, S. E., & Schiff, B. B. A biological theory of reinforcement. *Psychological Review*, 1967, **74**, 81–109.

Glickman, S. E., Higgins, T., & Isaacson, R. L. Some effects of hippocampal lesions on the behavior of Mongolian gerbils. *Physiology and Behavior*, 1970, **5**, 931–938.

Goodenough, F. L. Expression of the emotions in a blind-deaf child. *Journal of Abnormal Psychology*, 1932, **27**, 328–333.

Gottlieb, G., & Simner, M. L. Relationship between cardiac rate and nonnutritive sucking in human infants. *Journal of Comparative and Physiological Psychology*, 1966, **61**, 128–131.

Grandstaff, N. W. Frequency analysis of EEG during milk drinking. *Electroencephalography and Clinical Neurophysiology*, 1969, **27**, 57–65.

Grastyán, E., Lissák, K., Madarász, I., & Donhoffer, H. Hippocampal electrical activity during the development of conditioned reflexes. *Electroencephalography and Clinical Neurophysiology*, 1959, **11**, 409–430.

Green, J. D. The hippocampus. *Physiological Reviews*, 1964, **44**, 561–608.

Green, J. D., & Arduini, A. A. Hippocampal electrical activity in arousal. *Journal of Neurophysiology*, 1954, **17**, 533–557.

Green, J. D., Maxwell, D. S., Schindler, W. J., & Stumpf, C. Rabbit EEG "theta" rhythm: Its anatomical source and relation to activity in single neurons. *Journal of Neurophysiology,* 1960, **23**, 403–420.

Gross, C. G., Chorover, S. L., & Cohen, S. M. Caudate, cortical, hippocampal, and dorsal thalamic lesions in rats: Alternation and Hebb-Williams maze performance. *Neuropsychologia.* 1965, **3**, 53–68.

Hakerem, G., & Sutton, S. Pupillary response at visual threshold. *Nature,* 1966, **212**, 485–486.

Harper, R. Behavioral and electrophysiological studies of sleep and animal hypnosis. Unpublished doctoral dissertation, McMaster University, Hamilton, Ontario, 1968.

Hatch, A., Wiberg, G. S., Balazs, T., & Grice, H. C. Long-term isolation stress in rats. *Science,* 1963, **142**, 507.

Hebb, D. O. Drives and the CNS (conceptual nervous system). *Psychological Review,* 1955, **62**, 243–254.

Hebb, D. O. The distinction between "classical" and "instrumental." *Canadian Journal of Psychology,* 1956, **10**, 165–166.

Hebb, D. O. *A textbook of psychology.* Philadephia: Saunders, 1966.

Hebb, D. O., & Thompson, W. R. The social significance of animal studies. In G. Lindzey (Ed.), *Handbook of social psychology.* Cambridge, Mass.: Addison-Wesley, 1954.

Hodes, R., Peacock, S. M., & Heath, R. G. Influence of the forebrain on somatomotor activity. I. Inhibition. *Journal of Comparative Neurology,* 1951, **94**, 381–408.

Hoff, E. C., Kell, J. F., Jr., & Carrol, M. N., Jr. Effects of cortical stimulation and lesions on cardiovascular function. *Physiological Reviews,* 1963, **43**, 68–114.

Ingram, W. R., Barris, R. W., Fisher, C., & Ranson, S. W. Catalepsy: An experimental study. *American Medical Association Archives of Neurology and Psychiatry,* 1936, **35**, 1175–1197.

Ironside, R. Disorders of laughter due to brain lesions. *Brain,* 1956, **79**, 589–609.

Ito, M. Hippocampal electrical correlates of self-stimulation in the rat. *Electroencephalography and Clinical Neurophysiology,* 1966, **21**, 261–268.

Jackson, W. J. A comment on "The hippocampus and behavior." *Psychological Bulletin,* 1968, **69**, 20–22.

Jasper, H., & Penfield, W. Electrocorticograms in man: Effect of voluntary movement upon the electrical activity of the precentral gyrus. *Archiv für Psychiatrie und Zeitschrift Neurologie,* 1949, **183**, 163–174.

Kawamura, H., & Domino, E. F. Hippocampal slow ("arousal") wave activation in the rostral midbrain transected cat. *Electroencephalography and Clinical Neurophysiology,* 1968, **25**, 471–480.

Kawamura, H., Nakamura, Y., & Tokizane, T. Effect of acute brain stem lesions on the electrical activities of the limbic system and neocortex. *Japanese Journal of Physiology,* 1961, **11**, 564–575.

Kellogg, W. N. Communication and language in the home-raised chimpanzee. *Science,* 1968, **162**, 423–427.

Kenyon, J. The effect of septal lesions upon motivated behavior in the rat. Unpublished doctoral dissertation, McGill University, Montreal, Quebec, 1962.

Kimble, D. P. Possible inhibitory functions of the hippocampus. *Neuropsychologia,* 1969, **7**, 235–244.

Kimble, D. P., Rogers, L., & Hendrickson, C. W. Hippocampal lesions disrupt maternal, not sexual behavior in the albino rat. *Journal of Comparative and Physiological Psychology,* 1967, **63**, 401–407.

Kimble, G. A. *Hilgard and Marquis' conditioning and learning.* New York: Appleton-Century-Crofts, 1961.

Kimmell, H. D. Instrumental conditioning of autonomically mediated behavior. *Psychological Bulletin,* 1967, **67**, 337–345.

King, F. A. Effects of septal and amygdaloid lesions on emotional behavior and conditioned avoidance responses in the rat. *Journal of Nervous and Mental Disease,* 1958, **126**, 57–63.

Klemm, W. R. Electroencephalographic-behavioral dissociations during animal hypnosis. *Electroencephalography and Clinical Neurophysiology,* 1966, **21**, 365–372.

Klemm, W. R. Mechanisms of the immobility reflex ("animal hypnosis") II. EEG and multiple unit correlates in the brain stem. *Communications in Behavioral Biology,* 1969, **3**, 43–52.

Komisaruk, B. R. Synchrony between limbic system theta activity and rhythmical behavior in rats. *Journal of Comparative and Physiological Psychology,* 1970, **70**, 482–492.

Kramis, R. C., & Routtenberg, A. Rewarding brain stimulation, hippocampal activity, and foot-stomping in the gerbil. *Physiology and Behavior,* 1969, **4**, 7–11.

Krogh, A., & Lindhard, J. The regulation of respiration and circulation during the initial stages of muscular work. *Journal of Physiology,* 1913, **47**, 112–136.

Krogh, A., & Lindhard, J. A comparison between voluntary and electrically induced muscular work in man. *Journal of Physiology,* 1917, **51**, 182–201.

Lashley, K. S. The problem of serial order in behavior. In L. A. Jeffress (Ed.), *Cerebral mechanisms in behavior.* New York: Wiley, 1951.

Lewis, P. R., & Shute, C. C. D. The cholinergic limbic system: Projections to the hippocampal formation, medial cortex, nuclei of the ascending cholinergic reticular system, and the subfornical organ and supraoptic crest. *Brain,* 1967, **90**, 521–540.

Lhermitte, F., Gauthier, J. C., Marteau, R., & Chain, F. Troubles de la conscience et mutism akinetique. Étude anatomo-clinique d'un ramollissement paramédian, bilaterale du pédoncule cerebral et du thalamus. *Revue Neurologique,* 1963, **109,** 115–131.

Lindsley, D. B. Emotion. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951.

Lindsley, D. B., & Sassaman, W. H. Autonomic activity and brain potentials associated with "voluntary" control of the pilomotors (MM. arrectores pilorum). *Journal of Neurophysiology,* 1938, **1,** 342–349.

Lindsley, D. B., Schreiner, L. H., Knowles, W. B., & Magoun, H. W. Behavioral and EEG changes following chronic brain stem lesions in the cat. *Electroencephalography and Clinical Neurophysiology,* 1950, **2,** 483–498.

Lopes da Silva, F. H., & Kamp, A. Hippocampal theta frequency shifts and operant behavior. *Electroencephalography and Clinical Neurophysiology,* 1969, **26,** 133–143.

Low, L. A., & Low, H. I. Effects of CS-US interval length upon avoidance responding. *Journal of Comparative and Physiological Psychology,* 1962, **55,** 1059–1061.

MacDonnell, M. F., & Flynn, J. P. Control of sensory fields by stimulation of hypothalamus. *Science,* 1966, **152,** 1406–1408.

MacLean, P. D., Flanigan, S., Flynn, J. P., Kim, C., & Stevens, J. R. Hippocampal function: Tentative correlations of conditioning, EEG, drug, and radioautographic studies. *Yale Journal of Biology and Medicine,* 1955–1956, **28,** 380–395.

Magoun, H. W. Caudal and cephalic influences of the brain stem reticular formation. *Physiological Reviews,* 1950, **30,** 459–474.

Magoun, H. W. *The waking brain.* Springfield, Ill.: Charles C Thomas, 1963.

Maire, F. W. Cited in T. C. Ruch & H. D. Patton (Eds.), *Physiology and biophysics.* Philadelphia: Saunders, 1965, p. 501.

Malmo, R. B. Activation: A neuropsychological dimension. *Psychological Review,* 1959, **66,** 367–386.

Martin, J. P. *The basal ganglia and posture.* London: Pitman, 1967.

McBride, R. L., & Klemm, W. R. Mechanisms of the immobility reflex ("animal hypnosis") I. Influences of repetition of induction, restriction of auditory-visual input, and destruction of brain areas. *Communications in Behavioral Biology,* 1969, **3,** 33–41.

McCleary, R. A. Response specificity in the behavioral effects of limbic system lesions in the cat. *Journal of Comparative and Physiological Psychology,* 1961, **54,** 605–613.

McCleary, R. A. Response-modulating functions of the limbic system: Initiation and suppresion. *Progress in Physiological Psychology,* 1966, **1,** 209–272.

McNew, J. J., & Thompson, R. Role of the limbic system in active and passive avoidance conditioning in the rat. *Journal of Comparative and Physiological Psychology,* 1966, **61,** 173–180.

Miller, N. E. Central stimulation and other new approaches to motivation and reward. *American Psychologist,* 1958, **13,** 100–108.

Miller, N. E. Learning of visceral and glandular responses. *Science,* 1969, **163,** 434–445.

Milner, P. *Physiological psychology.* New York: Holt, Rinehart & Winston, 1970.

Mitchell, C. L., & Kaelber, W. W. Unilateral vs. bilateral medial thalamic lesions and reactivity to noxious stimuli. *American Medical Association Archives of Neurology,* 1967, **17,** 653–660.

Mogenson, G. J., & Peterson, R. J. Effects of spreading cortical depression on cardiac and somatomotor conditioned responses. *Canadian Journal of Physiology and Pharmacology,* 1966, **44,** 39–45.

Monnier, M., Hösli, L., & Krupp, P. Moderating and activating systems in medio-central thalamus and reticular formation. In R. Hernández Péon (Ed.), The physiological basis of mental activity. *Electroencephalography and Clinical Neurophysiology,* 1963, Suppl. No. 24, 97–112.

Mountcastle, V. B. *Medical physiology.* Vol. 1. Saint Louis: Mosby, 1968.

Nunnally, J. C., Knott, P. D., Duchnowski, A., & Parker, R. Pupillary response as a general measure of activation. *Perception and Psychophysics,* 1967, **2,** 149–155.

Obrist, P. Heart rate and somatic-motor coupling during classical aversive conditioning in humans. *Journal of Experimental Psychology,* 1968, **77,** 180–193.

Obrist, P. A., & Webb, R. A. Heart rate during conditioning in dogs: Relationship to somatic-motor activity. *Psychophysiology,* 1967, **4,** 7–34.

Obrist, P. A., Webb, R. A., & Sutterer, J. R. Heart rate and somatic changes during aversive conditioning and a simple reaction time test. *Psychophysiology,* 1969, **5,** 696–723.

Olds, M. E., & Olds, J. Approach-avoidance analysis of rat diencephalon. *Journal of Comparative Neurology,* 1963, **120,** 259–295.

Olton, D. S., & Isaacson, R. L. Effects of lateral and dorsomedial thalamic lesions on retention of active avoidance tasks. *Journal of Comparative and Physiological Psychology,* 1967, **64,** 256–261.

Paxinos, G., & Bindra, D. Rewarding intracranial stimulation, movement, and the hippocampal theta rhythm. *Physiology and Behavior,* 1970, **5,** 227–231.

Peacock, S. M., Jr., & Hodes, R. Influence of the forebrain on somatomotor activity: II. Facilitation. *Journal of Comparative Neurology,* 1951, **94,** 409–426.

Perez-Cruet, J., & Gantt, H. W. Relation between heart rate and "spontaneous" movements. *Bulletin of the Johns Hopkins Hospital,* 1959, **105,** 315–321.

Petsche, H., Stumpf, C., & Gogolák, G. The significance of the rabbit's septum as a relay station between the midbrain and the hippocampus. I. The control of hippocampal arousal activity by the septum cells. *Electroencephalography and Clinical Neurophysiology*, 1962, **14**, 202–211.

Pommereke, W. T. A study of the sensory areas eliciting the swallowing reflex. *American Journal of Physiology*, 1928, **84**, 36–41.

Pompeiano, O. The neurophysiological mechanisms of the postural and motor events during desynchronized sleep. *Research Publications, Association for Research in Nervous and Mental Disease*, 1967, **45**, 351–423.

Posluns, D. An analysis of chlorpromazine-induced suppression of the avoidance response. *Psychopharmacologia*, 1962, **3**, 361–373.

Posluns, D., & Vanderwolf, C. H. Improved avoidance performance following exploratory movement. *Canadian Journal of Psychology*, 1970, **24**, 499–504.

Radulovački, M., & Adey, W. R. The hippocampus and the orienting reflex. *Experimental Neurology*, 1965, **12**, 68–83.

Replogle, A. Motivational effects of lesions in the thalamus. Unpublished master's thesis, McGill University, 1960.

Rich, I., & Thompson, R. Role of the hippocampo-septal system, thalamus, and hypothalamus in avoidance conditioning. *Journal of Comparative and Physiological Psychology*, 1965, **59**, 66–72.

Roberts, L. Relationship of electrodermal and heart-rate responses to movement during discriminative classical conditioning in the rat. Paper presented at the meeting of the Eastern Psychological Association, Atlantic City, N. J., April 1970.

Roberts, W. W. Fear-like behavior elicited from dorsomedial thalamus of cat. *Journal of Comparative and Physiological Psychology*, 1962, **55**, 191–197.

Roberts, W. W., & Carey, R. J. Effect of dorsomedial thalamic lesions on fear in cats. *Journal of Comparative and Physiological Psychology*, 1963, **56**, 950–958.

Roth, S. R., Sterman, M. B., & Clemente, C. D. Comparison of EEG correlates of reinforcement, internal inhibition and sleep. *Electroencephalography and Clinical Neurophysiology*, 1967, **23**, 509–520.

Routtenberg, A., & Kramis, R. C. Hippocampal correlates of aversive midbrain stimulation. *Science*, 1968, **160**, 1363–1365.

Rushmer, R. F., & Smith, O. A., Jr. Cardiac control. *Physiological Reviews*, 1959, **39**, 41–68.

Sailer, S., & Stumpf, C. Beeinflussbarkeit der rhinencephalen Tatigkeit des Kaninchens. *Naunyn-Schmiedeberg's Archiv fur die Experimentelle Pathologie und Pharmakologie*, 1957, **231**, 63–77.

Sainsbury, R. S. Hippocampal activity during natural behavior in the guinea pig. *Physiology and Behavior*, 1970, **5**, 317–324.

Schrier, A. M. Effects of CS–US interval on avoidance conditioning of cats. *Journal of Genetic Psychology*, 1961, **98**, 203–210.

Schwartz, M. Conditioned stimulus variables in avoidance learning. *Journal of Experimental Psychology*, 1958, **55**, 347–351.

Sherrington, C. S. *The integrative action of the nervous system.* New Haven, Conn.: Yale University Press, 1961. (Originally published: London: Scribner, 1906).

Simner, M. L. Relationship between cardiac rate and vocal activity in newly hatched chicks. *Journal of Comparative and Physiological Psychology*, 1966, **61**, 496–498.

Simpson, H. M., & Paivio, A. Changes in pupil size during an imagery task without motor response involvement. *Psychonomic Science*, 1966, **5**, 405–406.

Simpson, H. M., & Paivio, A. Effects on pupil size of manual and verbal indicators of cognitive task fulfillment. *Perception and Psychophysics*, 1968, **3**, 185–190.

Smith, B. M. The polygraph. *Scientific American*, 1967, **216**, 25–31.

Smith, O. A., Jr., Rushmer, R. F., & Lasher, E. P. Similarity of cardiovascular responses to exercise and to diencephalic stimulation. *American Journal of Physiology*, 1960, **198**, 1139–1142.

Sommer-Smith, J. A., Powarzynski, J., Stirner, A., & Grünberg, V. Décharges cellulaires du noyau centre-médian du thalamus induites par la stimulation amygdalienne et septale. *Acta Neurologica Latinoamerica*, 1965, **11**, 360–367.

Storm van Leeuwen, W., Kamp, A., Kok, M. L., & Tielen, A. M. Relations between behavior in dogs and electrical activities in various parts of the brain. In E. A. Asratyan (Ed.), *Progress in brain research. Brain reflexes.* Amsterdam: Elsevier, 1967.

Stumpf, C. Drug action on the electrical activity of the hippocampus. *International Review of Neurobiology*, 1965, **8**, 77–138. (a)

Stumpf, C. The fast component in the electrical activity of the rabbits' hippocampus. *Electroencephalography and Clinical Neurophysiology*, 1965, **18**, 477–486. (b)

Taylor, J. *Selected writings of John Hughlings Jackson.* London: Staples Press, 1958.

Thompson, J. Development of facial expression in blind and seeing children. *Archives of Psychology*, 1941, **No. 264**, 1–47.

Thompson, R. Thalamic structures critical for retention of an avoidance conditioned response in rats. *Journal of Comparative and Physiological Psychology*, 1963, **56**, 261–267.

Tinbergen, N. *The study of instinct.* London: Oxford University Press, 1951.

Tizard, B., & Margerison, J. H. Psychological functions during wavespike discharge. *British Journal of Social and Clinical Psychology*, 1963, **3**, 6–15.

Tolman, E. C. Cognitive maps in rats and men. *Psychological Review*, 1948, **55**, 189–208.

Torii, S. Two types of pattern of hippocampal electrical activity induced by stimulation of hypothalamus and surrounding parts of rabbits' brain. *Japanese Journal of Physiology*, 1961, **11**, 147–157.

TORII, S., & KAWAMURA, H. Effects of amygdaloid stimulation on blood pressure and electrical activity of hippocampus. *Japanese Journal of Physiology*, 1960, **10**, 374–384.

UPRUS, V., GAYLOR, G. B., & CARMICHAEL, E. A. Shivering: A clinical study with especial references to the afferent and efferent pathways. *Brain*, 1935, **58**, 220–237.

VANDERWOLF, C. H. Medial thalamic functions in voluntary behaviour. *Canadian Journal of Psychology*, 1962, **16**, 318–330.

VANDERWOLF, C. H. The effect of medial thalamic lesions on previously established fear-motivated behaviour. *Canadian Journal of Psychology*, 1963, **17**, 183–187. (a)

VANDERWOLF, C. H. Improved shuttle-box performance following electroconvulsive shock. *Journal of Comparative and Physiological Psychlogy*, 1963, **56**, 983–986. (b)

VANDERWOLF, C. H. Effect of combined medial thalamic and septal lesions on active avoidance behavior. *Journal of Comparative and Physiological Psychology*, 1964, **58**, 31–37.

VANDERWOLF, C. H. Warm-up effects in the avoidance performance of rats with medial thalamic lesions. *Animal Behaviour*, 1966, **14**, 425–429.

VANDERWOLF, C. H. Behavioural correlates of "theta" waves. *Proceedings of the Canadian Federation of Biological Societies*, 1967, **10**, 41–42. (a)

VANDERWOLF, C. H. Effects of experimental diencephalic damage on food hoarding and shock avoidance behavior in the rat. *Physiology and Behavior*, 1967, **2**, 399–402. (b)

VANDERWOLF, C. H. Recovery from large medial thalamic lesions as a result of electroconvulsive therapy. *Journal of Neurology, Neurosurgery, and Psychiatry*, 1968, **31**, 67–72.

VANDERWOLF, C. H. Effects of chlorpromazine on avoidance behavior and on hippocampal electrical activity. Paper presented at the meeting of the Canadian Physiological Society, St. Adele, Quebec, January 1969. (a)

VANDERWOLF, C. H. Effects of medial thalamic damage on initiation of movement and learning. *Psychonomic Science*, 1969, **17**, 23–25. (b)

VANDERWOLF, C. H. Hippocampal electrical activity and voluntary movement in the rat. *Electroencephalography and Clinical Neurophysiology*, 1969, **26**, 407–418. (c)

VANDERWOLF, C. H., & HERON, W. Electroencephalographic waves with voluntary movement: Study in the rat. *American Medical Association Archives of Neurology and Psychiatry*, 1964, **11**, 379–384.

VANDERWOLF, C. H., & VANDERWART, M. L. Relations of heart rate to motor activity and arousal in the rat. *Canadian Journal of Psychology*, 1970, **24**, 434–441.

VANDERWOLF, C. H., & WHISHAW, I. Q. Is hippocampal electrical activity altered by training? Paper presented at the meeting of the Psychonomic Society, St. Louis, Missouri, November, 1969.

WAHLSTEN, D., COLE, M., SHARP, D., & FANTINO, E. Facilitation of bar-press avoidance by handling during the intertrial interval. *Journal of Comparative and Physiological Psychology*, 1968, **65**, 170–175.

WARREN, J. M., & AKERT, K. Impaired problem solving by cats with thalamic lesions. *Journal of Comparative and Physiological Psychology*, 1960, **53**, 207–211.

WHATMORE, G. B., & KOHLI, D. R. Dysponesis: A neurophysiologic factor in functional disorders. *Behavioral Science*, 1968, **13**, 102–124.

WHISHAW, I. Q. The influence of body temperature on hippocampal electrical activity in freely moving rats. Paper presented at the meeting of the Eastern Psychological Association, Atlantic City, N. J., April 1970.

WHISHAW, I. Q., & VANDERWOLF, C. H. Hippocampal correlates of movement. *Proceedings of the Canadian Federation of Biological Societies*, 1970, **13**, 48.

WILSON, S. A. K. Some problems in neurology. II. Pathological laughing and crying. *Journal of Neurology and Psychopathology*, 1924, **4**, 299–333.

WILSON, S. A. K. The narcolepsies. *Brain*, 1928, **51**, 63–109.

WISHART, T. The effects of lesions of the hippocampus, septum, and medial thalamus on a passive avoidance task. Unpublished master's thesis, University of Western Ontario, 1968.

WOODS, J. W. Behavior of chronic decerebrate rats. *Journal of Neurophysiology*, 1964, **27**, 634–644.

YOKOTA, T., & FUJIMORI, B. Effects of brainstem stimulation upon hippocampal electrical activity, somatomotor reflexes and autonomic functions. *Electroencephalography and Clinical Neurophysiology*, 1964, **16**, 375–383.

YOSHII, N., SHIMOKOCHI, M., MIYAMOTO, K., & ITO, M. Studies on the neural basis of behavior by continuous frequency analysis of EEG. In T. Tokizane & J. P. Schadé (Eds.), *Progress in brain research. Correlative neurosciences: Fundamental mechanisms.* Amsterdam: Elsevier, 1966.

YOUNG, G. A., & BLACK, A. H. The relationships between electrical activity of the hippocampus and overt skeletal behavior in the dog and rat. Paper presented at the meeting of the Eastern Psychological Association, Atlantic City, N. J., April 1970.

(Received June 2, 1970)

Psychological Review
1971, Vol. 78, No. 2, 114–121

VISUAL FEATURE-ANALYZERS AND AFTEREFFECTS OF TILT AND CURVATURE[1]

MAX COLTHEART[2]

University of Waterloo

The aftereffect produced by viewing a line tilted in the frontal plane can be explained in terms of the adaptation of orientation-specific units in the visual system. This explanation is more successful than explanations based on "normalization" or "satiation." Aftereffects of curvature or median-plane tilt can be explained in analogous ways, although the visual system units necessary are in this case mainly hypothetical, since experiments capable of demonstrating the existence of these types of units have not yet been performed.

Feature analyzers in the visual system of cats, monkeys, and humans have been extensively studied in recent years, both electrophysiologically (e.g., Hubel & Wiesel, 1962, 1965, 1968) and psychophysically (e.g., Campbell & Kulikowski, 1966; Gilinsky, 1968). Several authors (Barlow & Hill, 1963; Day, 1962, 1969; Dodwell & Gendreau, 1969; Mayo, Gilinsky, & Jochnowitz, 1968) have suggested that it should now be possible to explain several forms of aftereffect on the basis of these electrophysiological and psychophysical data. However, no detailed explanation of this sort has yet been published. The purpose of this paper is to develop an explanation of one aftereffect, the tilt aftereffect (Gibson & Radner, 1937), in terms of the orientation-specific contour detectors first described by Hubel and Wiesel (1962), and to evaluate the extent to which this explanation succeeds in dealing with what is known about the tilt aftereffect. Some comments will also be made concerning possible analogous explanations of other forms of aftereffect.

The relevant aspects of Hubel and Weisel's findings are, briefly, as follows: (*a*) There exist units in the visual cortex of the cat and the monkey which respond to straight-line contours (edges, light slits, or dark bars); (*b*) each unit has a "preferred orientation," that is, it responds to one particular orientation of a contour. Different units have different preferred orientations; (*c*) any such unit exhibits a decline in response rate with protracted stimulation. After such stimulation, the unit is in a state of reduced sensitivity, from which it recovers over time, gradually regaining normal sensitivity; and (*d*) although a unit responds best to a contour having the appropriate orientation (i.e., the unit's "preferred orientation"), it will also respond, though less strongly, to contours with somewhat different orientations. The strength of the unit's response is inversely related to the difference between the unit's preferred orientation and the orientation of the stimulus. No response occurs if this difference exceeds a certain value: ±10° to 15° in the cat, ±5° to 10° in the monkey (Hubel & Wiesel, 1968, p. 220). The corresponding value for humans is uncertain (as is discussed below), but all that is required for the purposes of this paper is that response strength should decline monotonically with the absolute discrepancy between the unit's preferred orientation and the orientation of the stimulus.

It follows from these data that when a straight-line contour is being inspected, a range of orientation-sensitive units will be stimulated. Those units whose preferred orientation coincides with the orientation of the stimulus will respond most strongly. Other units will respond less strongly, their response rate depending inversely on the degree of departure of the stimulus orientation from the unit's preferred orientation. We suppose that the perceived orientation yielded by this pattern of response is given by averaging the preferred orientations of all the units which respond, weighting each of these preferred orientations by the extent to which the unit is responding above its maintained discharge level.

---

[1] The preparation of this paper was supported by Grant No. APA-0350 from the National Research Council of Canada. I am indebted to Charles S. Harris for various suggestions and for criticisms of an earlier draft.

[2] Requests for reprints should be sent to Max Coltheart, Department of Psychology, University of Waterloo, Ontario, Canada.

Let us assume, for expository purposes, that for humans the range of orientations to which a single unit will respond is ±20° about its preferred orientation (as suggested by Campbell & Kulikowski, 1966, and Mayo et al., 1968). Now, suppose a subject inspects an I figure, a line tilted at +10° (the plus sign indicating a clockwise departure from 0°, the vertical). In this case, units with preferred orientations between −10° and +30° will respond to the +10° I figure. With protracted viewing, these units will adapt. If a vertical line, the T figure, is now inspected, units from −20° to +20° will be stimulated. Units with preferred orientations from −20° to −10° will respond with normal vigor, since they were not stimulated during inspection of the I figure, and hence will not have adapted. However, units with preferred orientations from −10° to +20° will be in a state of reduced sensitivity because they were stimulated throughout the inspection period. They will respond less strongly therefore to the vertical stimulus than they ought. As a consequence, the weighted average response to the vertical line will be biased in the negative direction, instead of being 0°, as would normally be the case. Therefore, the vertical line will be perceived as having a counterclockwise tilt; that is, there will be a negative aftereffect.

COMPARISON WITH OTHER THEORIES

There are two major theories of the tilt aftereffect: the normalization theory (Gibson & Radner, 1937) and the satiation theory (Kohler & Wallach, 1944). According to the normalization theory, the vertical and the horizontal are norms of visual space, and the effect of viewing a line which departs somewhat from one of these norms is to shift the norm in the direction of the inspected line. A line which formerly coincided with this norm will now appear to depart from it in a direction opposite to the direction of tilt of the I figure. Thus a negative aftereffect will occur.

Difficulties with this theory have been pointed out by Morant and Harris (1965). The main difficulty is that inspection of an I figure which coincides with a norm (i.e., is vertical or horizontal) generates an aftereffect (Kohler & Wallach, 1944, pp. 309–311; confirmed by Templeton, Howard, & Easting, 1965). This is clearly inconsistent with the normalization theory because this kind of I figure could not distort the norms of visual space. On the other hand, this effect would be expected to occur if the tilt aftereffect is a consequence of the adaptation of a range of orientation-specific units, because this adaptation would be produced by I figures of *any* orientation.

According to the satiation theory, protracted inspection of a visual stimulus leads to "satiation" of the area of the visual cortex to which this stimulus is projected. Stimuli subsequently projected to adjacent cortical areas will be displaced from the satiated area. This theory, too, suffers from an insuperable difficulty: the "indirect effect." Viewing a line tilted slightly clockwise from the vertical will make an objectively vertical line appear to have anticlockwise tilt (the "direct effect"); it will also make an objectively horizontal line appear to be slightly tilted counterclockwise from the horizontal (the "indirect effect"). This indirect effect, discovered by Gibson and Radner (1937) and verified by Morant and Harris (1965), cannot be explained by the satiation theory.

The indirect effect can be understood in terms of the analysis being proposed in this paper, by virtue of a property exhibited by one class of orientation-sensitive unit, the "higher order hypercomplex cell" (Hubel & Wiesel, 1965, 1968). This type of unit has *two* preferred orientations, 90° apart. Viewing a +10° line would adapt such units with preferred orientations from −10° to +30° and +80° to +120°. Now, when a +90° line is viewed, the adapted state of the +80° to +120° units will bias the weighted average response in an anticlockwise direction, thereby yielding an "indirect effect."

Clearly, units of this type cannot be the only units involved in tilt detection; otherwise we could not discriminate 0° from 90°. Simple cells, complex cells, and lower order hypercomplex cells, which all have single preferred orientations, must also be involved. This means that the adaptation of several classes of cell contributes to the direct effect, whereas the indirect effect is produced by the adaptation of only one of these classes. It is not surprising, therefore, that the indirect effect is smaller than the direct effect, a result which embarrasses the normalization theory. Gibson attributed the indirect effect to a hypothetical linkage between the vertical and horizontal norms. If the norms are linked, then if one is displaced, the other must be too. He then suggested that the indirect effect is smaller than the direct because the linkage between the two norms is an insecure one, there being "a certain amount of 'play' or 'lost motion' between them [Gibson & Radner, 1937, p. 464]." This explanation is patently ad hoc.

Thus neither the normalization theory nor the satiation theory is consistent with what is known about the tilt aftereffect, whereas an analysis of this aftereffect in terms of the adaptation of orientation-specific units in the visual system is capable of encompassing the major features of the tilt aftereffect. The direct relationship between aftereffect magnitude and duration of inspection of the I figure, and the inverse relationship between aftereffect magnitude and time since I-figure offset, are obviously deducible from this analysis. So is the "distance paradox": In general, the magnitude of the tilt aftereffect is inversely related to the angular separation between the I figure and the T figure, except when the two figures have almost identical orientations. Here the aftereffect is very small, and an increase in the angular separation of the two figures can produce a larger aftereffect. This is to be expected because when the I figure and T figure almost coincide, the distribution of adapted units is almost symmetrical about the T figure. Only when the two figures differ appreciably in orientation will a sizable asymmetry begin to emerge.

The general function relating the magnitude of the tilt aftereffect to the difference between T-figure orientation and I-figure orientation has been described in detail (e.g., Morant & Harris, 1965). If the selectivity of orientation-specific units (i.e., the range of orientations capable of stimulating a given unit) were known, then it would be possible to predict the exact form of the angular function of the tilt aftereffect, rather than merely predicting its curvilinearity. This would be a powerful test of any theory attempting to explain the tilt aftereffect in terms of the adaptation of orientation-specific units. Unfortunately, precise physiological measures of this selectivity in humans are as yet unavailable, and attempts to estimate it from psychophysical data have yielded very diverse figures.[3] Consequently it would be premature to attempt to predict the function relating aftereffect magnitude to I-figure orientation using data from experiments directly investigating orientational selectivity.

TWO DIFFICULTIES

An analysis of the tilt aftereffect in terms of the adaptation of orientation-specific units appears to offer considerable promise. Difficulties for such an analysis are raised by the distinction between adaptation and aftereffect, and also by Day and Wade's (1969) conclusion that it is orientation with respect to gravity, not to the retina, that is crucial for the tilt aftereffect.

---

[3] Parlee (1969), using single lines for both test and adapting stimuli, found that even when the two lines differed in orientation by 90°, some contour-specific masking occurred. Sekuler (1965), using a single line as the test stimulus and a grid as the adapting stimulus, found orientation-related masking to be present when the angular separation of the two stimuli was 45°, but not when it was 60°. Campbell and Kulikowski (1966), using grids for both test and adapting stimuli, found no masking effect when the stimuli differed by more than 20° in orientation; Mayo et al. (1968), also using grids for test and adapting stimuli, obtained a value between 15° and 20° for the limit of the masking effect. One could attempt to reconcile the discrepancies among these data by suggesting that a grid of lines produces a much more narrowly tuned response than does a single line, although there is no obvious physiological evidence for such an assumption. If this is correct, grid stimuli should produce a smaller tilt aftereffect than line stimuli.

Two sorts of effect are noted during experimental studies of the tilt aftereffect. An *aftereffect* is manifested during the postinspection presentation of the T figure, and is measured by a comparison of preinspection and postinspection responses to the T figure. *Adaptation* (which will be called "neutralization," following Taylor, 1962, in view of the variety of uses of the term "adaptation") is manifested during the inspection period by a progressive change over time in the appearance of the I figure, which appears to approach verticality if it was originally close to the vertical, or horizontality if it was originally close to the horizontal (Gibson, 1934; Gibson & Radner, 1937; Morant & Mistovich, 1960; Vernon, 1934).

It might seem obvious that neutralization and aftereffect are merely two facets of a single process, as Gibson originally assumed; he still holds this view (Gibson, 1966, p. 309). However the analysis of the tilt aftereffect in terms of orientation-specific units applies only to aftereffects, not to neutralization. There is no way to make such an analysis predict changes in the appearance of the I figure while it is being continuously observed. The I figure will adapt units equally on both sides of its own orientation, so no bias will be introduced. This does not indicate that the proposed analysis is deficient, however, because the aftereffect and neutralization may well be quite separate effects. They behave differently in several ways; for example, an I figure of any orientation produces an aftereffect, whereas vertical, horizontal, or 45° I figures do not produce neutralization. Second, neutralization is not retinally localized, whereas the magnitude of the aftereffect declines markedly as the retinal location of the I and T figures are moved apart (Morant & Mikaelian, 1960).[4] Consequently, as Taylor (1962) has pointed out, the two phenomena require different explanations, and a single theory cannot explain both. Indeed, Morant and Harris (1965) concluded that tilt experiments involve two underlying processes, satiation and normalization. They suggested that neutralization and the "indirect effect" reflect normalization, whereas the aftereffect is typically the algebraic sum of normalization and satiation effects.[5]

It is only the aftereffect which is here attributed to the adaptation of orientation-specific units. A possible neurophysiological mechanism for the neutralization effect remains to be discovered. One such mechanism might involve units which signal, not orientation, but departure from the vertical or horizontal. If response rate were proportional to departure from the "norm," and if adaptation reduced response rate, then neutralization effects would occur during prolonged viewing. Whether this type of unit contributes to the aftereffect, or is merely responsible for neutralization, remains to be determined. Templeton et al. (1965) found that the aftereffect obtained with a vertical I figure was significantly smaller than the aftereffect obtained with a +10° I figure. Differences between their two conditions make this comparison a dubious one, but if the result is accepted, it might be interpreted as being due to the fact that when the I figure is vertical, only orientation-specific units are adapted, whereas both orientation-specific units and "norm-deviation" units are adapted by an oblique I figure. This interpretation is quite consistent with the view that the mechanisms underlying neutralization are not identical with the mechanisms underlying the aftereffect.

---

[4] Since this was written, Muir and Over (1970) have reported a failure to replicate Morant and Mikaelian's results.

[5] It might be suggested that differences in the properties of neutralization and aftereffect could be explained in terms of differences in the sensitivity curves of units having different preferred orientations. Such differences were postulated by Andrews (1967), who inferred from his data that units maximally sensitive to vertical, horizontal, or 45° contours have symmetrical sensitivity distributions, whereas other orientation-specific units have skewed sensitivity distributions. Perhaps, then, units with symmetrical sensitivity distributions would not, while adapting out, lead to a drift of apparent orientation, but would still produce an aftereffect. How this unitary explanation could be worked out in detail is not evident, and at any rate it fails to account for the fact that neutralization is not retinally localized, whereas the aftereffect is (Morant & Mikaelian, 1960).

The second difficulty for this analysis is raised by the work of Day and Wade (1969) who used body tilt in order to separate retinal and gravitational orientation. They found a tilt aftereffect when the I and T figures were retinally identical, but differed in gravitational orientation. This finding poses a problem for any analysis that is based on units sensitive to retinal orientation, as is considered to be the case for the type of unit studied by Hubel and Wiesel (1962, 1965, 1968). Fortunately, recent physiological evidence (Horn & Hill, 1969; Spinelli, 1970) has revealed that there are visual-system units which respond to particular gravitational orientations. In these experiments, receptive-field orientation in the cat was studied during body tilt. For a small number of orientation-specific units, the receptive-field axis did not tilt as much as the cat's retina did. In fact, some units exhibited almost no change in receptive-field axis (relative to the environment) with a head tilt of 30°; that is, these units were responding almost entirely to gravitational orientation rather than retinal orientation.

There seems to be no doubt that a difference between gravitational orientations of I and T figures is *sufficient* to generate an aftereffect, which could be mediated by the gravitation-sensitive units mentioned above. However, Day and Wade also argued that a difference in gravitational orientation is *necessary* for a tilt aftereffect, since they found no aftereffect when the I and T figures differed in retinal orientation but had the same gravitational orientation. This second conclusion is difficult to accept, and further work is required. Day and Wade did obtain aftereffects in the predicted direction in two experiments when only retinal orientation was varied (though the effects were not significant); moreover, the aftereffect was much greater when both retinal and gravitational orientation were varied than when only the latter was varied. Moreover, as they mention, ocular countertorsion (which occurs when the body is tilted) would enhance any aftereffect attributable to gravitational orientation and reduce any aftereffect attributable to retinal orientation. Clearly this is an issue which needs further investigation. The existence of an aftereffect due solely to differences in gravitational orientation, at any rate, is not inconsistent with the view that tilt aftereffects are due to the adaptation of orientation-specific units, since some such units are responsive to gravitational orientation.

OTHER FORMS OF AFTEREFFECT

There are two other aftereffects whose properties are similar to those of the tilt aftereffect:[6] the curvature aftereffect (Carlson, 1963; Gibson, 1933) and the "third-dimensional" or median-plane tilt aftereffect (Bergman & Gibson, 1959; Kohler & Emery, 1947; Wenderoth, 1967; Wenderoth, Rodger, & Curthoys, 1968). The resemblances between these aftereffects and the tilt aftereffect suggest that it might be fruitful to attempt to analyze them in the same way as the tilt aftereffect.

Prolonged inspection of a curved line makes a straight line appear to be curved in the opposite direction. This hints at the existence of units sensitive to direction or degree of curvature. For example, an analysis of the curvature aftereffect precisely analogous to that given for the tilt aftereffect could be developed starting from the postulate that units exist which respond best to contours having a particular degree of curvature, but which also respond (though less strongly) to contours whose curvature approximates the preferred curvature. One prediction entailed by this analysis is that an extremely curved I figure should produce a smaller curvature aftereffect than an I figure of lesser curvature, assuming a straight-line T figure. This was found by Bales and Follansbee (1935).

It is not difficult to imagine how a set of orientation-specific cells could be wired up to a higher order unit in such a way as to produce a unit maximally sensitive to contours having a particular degree of curvature. In fact Hubel and Wiesel have occasionally made guarded references to an interpretation of some of the units they have found as curvature detectors (Hubel & Wiesel, 1965, p. 286; 1968, p. 238). Furthermore, one class of unit described in the frog by Lettvin, Maturana, Pitts, and McCulloch (1961) has as its preferred stimulus a curved contour.

---

[6] Blakemore and Sutton (1969) have suggested that a different form of aftereffect, the displacement aftereffect (Kohler & Wallach, 1944), might be explicable in terms of the adaptation of size-specific units.

The median-plane tilt aftereffect occurs after inspection of a line tilted toward or away from the observer in his median plane; an objectively vertical line will now appear tilted in the median plane in the opposite direction. An analogous explanation of this aftereffect would require the existence of units sensitive to tilt in the median plane and with a preferred median-plane orientation. Such units have not been observed; but no one has looked for this particular type of unit, and its feasibility is indicated by the work of Barlow, Blakemore, and Pettigrew (1967) and Pettigrew, Nikara, and Bishop (1968), who showed that different binocularly driven cortical units have different preferred receptive-field disparities. This means that with convergence fixed, different units are maximally excited by stimuli at different distances. If a number of such units had receptive fields arranged in a vertical sequence on both retinae, and if the preferred disparity of the units increased progressively as a function of their vertical retinal coordinate, this aggregation of units would be maximally stimulated by a line having a particular median-plane tilt. A higher order unit on which this set of units converged would have a preferred median-plane tilt and would serve therefore as the mechanism underlying the median-plane tilt aftereffect.

The distinction between neutralization and aftereffect is important for the curvature and median-plane tilt aftereffects as well as for the frontal-plane tilt aftereffect. During inspection, a curved line appears to become less curved (Gibson, 1933), and a line tilted from the vertical in the median plane appears to approach verticality (Bergman & Gibson, 1959). In terms of the present analysis, these two forms of neutralization suggest the existence of units whose response rate is proportional to the degree of departure from linearity of a curved line and to the degree of departure from verticality of a line tilted in the median plane.[7]

CONCLUSIONS

This paper has proposed explanations of various types of aftereffect in terms of the adaptation of various types of feature-analyzing units. Whether such explanations are of value has been questioned by Weisstein (1969), who has claimed that perceptual aftereffects are useless as a paradigm for studying the characteristics of visual feature-analyzers. Her two main arguments are that (*a*) for the motion aftereffect, at least, a heterogeneous collection of units seems to underlie the aftereffect, and the aftereffect paradigm cannot distinguish among the different classes of unit; and (*b*) orientation-specific color aftereffects, which were assumed to be direct evidence for chromatic adaptation of edge detectors (McCollough, 1965), can also be explained in terms of a simpler kind of neural unit, the "dipole" (Harris & Gibson, 1968) and so, again, the aftereffect paradigm fails to give unambiguous information about neural circuitry.

However, both of these arguments are irrelevant to the present paper because no attempt is being made to use aftereffect data to demonstrate the existence of particular neural circuits. Instead, where something is known about the circuitry, the aim is to show that the properties of aftereffects can be deduced from the circuits. Where nothing is known about the circuitry, the aim is to suggest the characteristics of the circuits underlying aftereffects. Direct investigations of these underlying mechanisms would then be required if this explanation of the aftereffect is to be pursued. In neither instance is it being assumed that aftereffect data provide a direct route to knowledge of the underlying neural circuits.

---

[7] Since this was written, Fiorentini and Maffei (1970) have reported that the amplitude of the evoked potential in humans viewing a pair of stereoscopically presented gratings is correlated with the degree of perceived inclination of the fused image. It could be argued that their measurements reflect the activity of units whose response rate is proportional to the departure from zero tilt of a stimulus tilted in the horizontal plane. Except for the substitution of the horizontal for the median plane, these are the units postulated above. Such units should exist for both types of tilt since neutralization occurs in both planes. It would be of the greatest interest to record the amplitude of these evoked potentials during the neutralization which would occur during prolonged inspection.

As far as the median-plane tilt aftereffect and the curvature aftereffect are concerned, all that has been done in this paper is to show that plausible neural units can be designed which would produce such aftereffects. What is required next is independent evidence for the existence of such units. A variety of techniques is available by means of which such evidence might be gathered. Direct electrophysiological studies, or psychophysical studies using such paradigms as those employed by Gilinsky (1968) or McCollough (1965), might be of value, but the most promising approach would appear to be the evoked-potential technique used by Campbell and Maffei (1970) and Fiorentini and Maffei (1970).

As far as the frontal-plane tilt aftereffect is concerned, it appears that present physiological knowledge is sufficient to provide an explanation for this aftereffect, although the neutralization effect and the possible importance of gravitational orientation raise questions which remain to be answered.

REFERENCES

Andrews, D. P. Perception of contour orientation in the central fovea. Part I: Short lines. *Vision Research*, 1967, **7**, 975–997.

Bales, J. F., & Follansbee, G. L. The aftereffect of the perception of curved lines. *Journal of Experimental Psychology*, 1935, **18**, 499–503.

Barlow, H. B., Blakemore, C., & Pettigrew, J. D. The neural mechanism of binocular depth discrimination. *Journal of Physiology*, 1967, **193**, 327–342.

Barlow, H. B., & Hill, R. M. Evidence for a physiological explanation of the waterfall phenomenon and figural aftereffects. *Nature*, 1963, **200**, 1345–1347.

Bergman, R., & Gibson, J. J. The negative aftereffect of the perception of a surface slanted in the third dimension. *American Journal of Psychology*, 1959, **72**, 363–374.

Blakemore, C., & Sutton, P. Size adaptation: A new aftereffect. *Science*, 1969, **166**, 245–247.

Campbell, F. W., & Kulikowski, J. J. Orientational selectivity of the human visual system. *Journal of Physiology*, 1966, **187**, 437–445.

Campbell, F. W., & Maffei, L. Electrophysiological evidence for the existence of orientation and size detectors in the human visual system. *Journal of Physiology*, 1970, **207**, 635–652.

Carlson, V. R. The generality of negative aftereffect following adaptation to curvature. *Scandinavian Journal of Psychology*, 1963, **4**, 129–133.

Day, R. H. Excitatory and inhibitory processes as the basis of contour shift and negative aftereffect. *Psychologia*, 1962, **5**, 185–193.

Day, R. H. *Human perception*. Sydney: Wiley, 1969.

Day, R. H., & Wade, N. J. The reference for visual normalization. *American Journal of Psychology*, 1969, **82**, 191–197.

Dodwell, P. C., & Gendreau, L. Figural aftereffects, sensory coding, expectation and experience. *British Journal of Psychology*, 1969, **60**, 149–167.

Fiorentini, A., & Maffei, L. Electrophysiological evidence for binocular disparity detectors in human visual system. *Science*, 1970, **169**, 208–209.

Gibson, J. J. Adaptation aftereffect and contrast in the perception of curved lines. *Journal of Experimental Psychology*, 1933, **16**, 1–31.

Gibson, J. J. Vertical and horizontal orientation in visual perception. *Psychological Bulletin*, 1934, **31**, 739. (Abstract)

Gibson, J. J. *The senses considered as perceptual systems*. Boston: Houghton Mifflin, 1966.

Gibson, J. J., & Radner, M. Adaptation, aftereffect and contrast in the perception of tilted lines. I. Quantitative studies. *Journal of Experimental Psychology*, 1937, **20**, 453–467.

Gilinsky, A. S. Orientation-specific effects of patterns of adapting light on visual acuity. *Journal of the Optical Society of America*, 1968, **58**, 13–18.

Harris, C. S., & Gibson, A. R. Is orientation-specific color adaptation in human vision due to edge detectors, afterimages or "dipoles"? *Science*, 1968, **162**, 1506–1507.

Horn, G., & Hill, R. M. Modifications of receptive fields in the visual cortex occurring spontaneously and associated with bodily tilt. *Nature*, 1969, **221**, 186–188.

Hubel, D. H., & Wiesel, T. N. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Journal of Physiology*, 1962, **160**, 106–154.

Hubel, D. H., & Wiesel, T. N. Receptive fields and functional architecture in two non-striate visual areas (18 and 19) of the cat. *Journal of Neurophysiology*, 1965, **28**, 229–289.

Hubel, D. H., & Wiesel, T. N. Receptive fields and functional architecture of monkey striate cortex. *Journal of Physiology*, 1968, **195**, 215–243.

Kohler, W., & Emery, D. A. Figural aftereffects in the third dimension of visual space. *American Journal of Psychology*, 1947, **60**, 159–201.

Kohler, W., & Wallach, H. Figural aftereffects: An investigation of visual processes. *Proceedings of the American Philosophical Society*, 1944, **88**, 269–357.

Lettvin, J. A., Maturana, H. R., Pitts, W. H., & McCulloch, W. S. Two remarks on the visual system of the frog. In W. A. Rosenblith (Ed.), *Sensory communication*. Cambridge: M.I.T. Press, 1961.

Mayo, T. H., Gilinsky, A. S., & Jochnowitz, A. Visual masking and tilt aftereffects as a function of tilt and angular separation between test and adaptation line patterns. *Proceedings of the 76th Annual Convention of the American Psychological Association*, 1968, **3**, 97–98.

McCollough, C. Color adaptation of edge-detectors in the human visual system. *Science*, 1965, **149**, 1115–1116.

Morant, R. B., & Harris, J. R. Two different aftereffects of exposure to visual tilts. *American Journal of Psychology*, 1965, **78**, 218–226.

Morant, R. B., & Mikaelian, H. Inter-field tilt aftereffects. *Perceptual and Motor Skills*, 1960, **10**, 95–98.

Morant, R. B., & Mistovich, M. Tilt aftereffects between the vertical and horizontal axes. *Perceptual and Motor Skills*, 1960, **10**, 75–81.

Muir, D., & Over, R. Tilt aftereffects in central and peripheral vision. *Journal of Experimental Psychology*, 1970, **85**, 165–170.

Parlee, M. B. Visual backward masking of a single line by a single line. *Vision Research*, 1969, **9**, 199–205.

Pettigrew, J. D., Nikara, T., & Bishop, P. O. Binocular interaction on single units in cat striate cortex: Simultaneous stimulation by single moving slit with receptive fields in correspondence. *Experimental Brain Research*, 1968, **6**, 391–410.

Sekuler, R. W. Spatial and temporal determinants of visual backward masking. *Journal of Experimental Psychology*, 1965, **70**, 401–406.

Spinelli, D. N. Recognition of visual patterns. In D. A. Hamburg (Ed.), *Perception and its disorders*. Baltimore: Williams & Wilkins, 1970.

Taylor, M. M. Figural after-effects: A theory. *Canadian Journal of Psychology*, 1962, **16**, 247–277.

Templeton, W. B., Howard, I. P., & Easting, G. Satiation and the tilt aftereffect. *American Journal of Psychology*, 1965, **78**, 656–659.

Vernon, M. D. The perception of inclined lines. *British Journal of Psychology*, 1934, **25**, 186–196.

Weisstein, N. What the frog's eye tells the human brain: Single cell analyzers in the human visual system. *Psychological Bulletin*, 1969, **72**, 157–176.

Wenderoth, P. M. The angular function of an aftereffect of slant: A methodological study. Unpublished master's thesis, University of Sydney, 1967.

Wenderoth, P. M., Rodger, R. S., & Curthoys, I. S. Confounding of psychophysical errors and sensory effects in adjustment measures of spatial aftereffects. *Perception and Psychophysics*, 1968, **4**, 133–138.

(Received May 19, 1970)

Psychological Review
1971, Vol. 78, No. 2, 122–129

PROBLEM OF SERIAL ORDER IN BEHAVIOR IS NOT RESOLVED BY CONTEXT-SENSITIVE ASSOCIATIVE MEMORY MODELS[1]

TERRY HALWES
University of Connecticut
and Haskins Laboratories

AND

JAMES J. JENKINS[2]
University of Minnesota

Associative-chain theories have trouble accounting for the basic fact that the elements of behavior are ordered serially. Wickelgren attempted to avoid these difficulties by postulating a "context-sensitive associative theory," "which assumes that serial order is encoded by means of associations between context-sensitive elementary motor responses." This reply argues that (*a*) such theories are not adequate to account for serial ordering and the other facts of speech production, (*b*) that arguments advanced in their support are incorrect, and (*c*) the nonassociative alternative considered and rejected by Wickelgren is in no way related to seriously held nonassociative theories.

In an article aimed at countering Lashley's (1951) paper on the problem of serial order in behavior, Wickelgren (1969) proposed to save some form of associative theory by taking units at a basic, "noncreative," elementary motor response level and allowing them to combine in "context-sensitive associations." He illustrated his thesis with the case of speech production, choosing the phoneme as the individual unit and the ordering of phonemes into words as the relevant problem of serial order. He argued in general that such models are sufficient to account for the problem of serial order in all "noncreative" behavior.

The writers agree with Wickelgren's contention that there is no extant model that can pretend to cope with even the most obvious facts presently known about speech production or any of the other kinds of human skills that he discusses. Further, we wholeheartedly agree that attempts are desperately needed to fill this void. Finally, we applaud Wickelgren's step toward greater freedom in the kind of model he employs in approaching the difficult problems involved in dealing with temporal integration. Having made these approving remarks, we must nevertheless criticize Wickelgren's theoretical formulation as being severely limited, and must point out its inability to cope with the phenomena of speech production toward which it was directed. Specifically, we argue that (*a*) Wickelgren has misunderstood Lashley's basic position; (*b*) Wickelgren's distinction between creative and noncreative behavior will engender serious misunderstandings of the notions involved in generative conceptions of behavior; and (*c*) the models Wickelgren has proposed are scarcely more powerful or adequate than the associative models already on hand.

We are not, of course, in a position to propose explicit alternative models which could conceivably be correct. Were we so fortunate, there would be no need for this paper. We have no such candidate because we have not yet been able to devise a theory that seemed capable of accounting for certain facts that *must* be dealt with by any theory of skilled motor acts. The need to reemphasize these facts is what prompts the present paper.

---

[1] Preparation of this paper was supported in part by grants to the University of Minnesota, Center for Research in Human Learning, from the National Science Foundation (GB-17590), the National Institute of Child Health and Human Development (HD-01136), and the Graduate School of the University of Minnesota. We appreciate the efforts of James Tweedy and Walter Weimer in trying to keep us from error. They are, of course, not to blame for any failure in their task.

[2] Requests for reprints should be sent to James J. Jenkins, 400 Ford Hall, University of Minnesota, Minneapolis, Minnesota 55455.

Comment on Lashley's Analysis

Rereading Lashley's (1951) paper suggests to us that Wickelgren's first assumption is a consequence of mistaking a pedagogical device for a theoretical position. Lashley's major intent was to call attention to the failure of association theory in the hands of both the Structuralists and the Behaviorists. He was convinced by his physiological and behavioral work that the nervous system was dynamic and organized in highly complex systems. In his article, he suggested behavioral examples that would serve as convincing indicators of these complexly organized systems. As Wickelgren points out, Lashley's examples were characterized in terms of "unitary acts" which were not themselves considered as containing any temporal relationships. He used this simplification to make the point that the ordering of those units was a complex problem involving at least *determining tendencies* and the *syntax* of the act. We do not believe that he seriously intended these unitary acts as a "rock bottom" elementary response level. At one point in the article, he talks of the individual movements of the speech articulators; at another, of the individual production of phonemes; and at another, of the activation of individual words. Throughout the article, it is clear that he was thinking of schemata and of hierarchies of organization prevading all skilled acts "even down to the temporal coordination of muscular contractions in such a movement as reaching and grasping [p. 123]." It seems unlikely to us that he would have been impressed by an associative analysis of any arbitrary level (such as the production of phonemes). We suspect that he would have pointed out that this level, too, is not composed of simple units, and that the whole problem could be restated in terms of further "unitary responses" below these complex phones.

Creative and Noncreative Behavior

Wickelgren (1969) has proposed that behavior be separated into two broad classes:

> It seems quite useful to distinguish behavior which is fundamentally creative, though lawful, such as speech at the level of syntax, from behavior which is fundamentally noncreative, such as speech at the level of word pronunciation. . . . We say the same words over and over again. We rarely say a sentence identical to one said before. This seems to me, as it has to many other people concerned with language (Chomsky, 1964; Chomsky & Miller, 1963), to be perhaps the most fundamental fact about language at the level of syntax. If one feels this to be such a fundamental feature of syntax, one should certainly not expect to account for the serial order of phonemes in words in exactly the same manner as one accounts for the serial order of words (or morphemes) in sentences. Thus, it seems important to distinguish creative and noncreative serially ordered behavior in language and in all other behavior [p. 2].

Two aspects of this must be discussed. First, the logic is misleading. Surely no one intends to account for phonology "in exactly the same manner as one accounts for the serial order of words." We confess that we do not even know what that might mean. But this argument must not prevent our considering the possibility that phonology may be as complex as syntax.

Second, and much more important, this distinction seems to us to represent a fundamental misunderstanding of the notion of "creativity of behavior" as elaborated, for example, by Chomsky and his colleagues (Chomsky, 1964; Chomsky & Miller, 1963). Wickelgren does not provide us with a statement of what he means by the term "creative," but it seems fair from his examples to infer that this concept is closely linked to the notion of infinite or potentially infinite variation. Although this argument has often been presented to overwhelm simple models of language production, and although one often hears the problem of creativity characterized as the problem of how the organism makes "infinite use of finite means" (Chomsky, 1966), this simple characterization, unless qualified and elaborated, may be very misleading. Impossibility proofs are easy ways to defeat alternative models which can not produce infinitely many sentences o sentences of infinite length. But impossibility proofs do not, in themselves, constitut the evidence supporting the claim that mos human (and, perhaps, animal) behavior fundamentally creative. Such proofs requi assumptions that no one wants to pu very far. No *real* sentence ever was or ever will be of unlimited length, either temporally or in cardinality of words or phonemes; it is equally clear that there is not an unlimited number of words in a given language at a given time. The important consideration is not that behavior is creative because it comes in (potentially) an infinite number of different versions but, rather, that any reasonable theory of sentence production must recognize that the limitations on the length and number of sentences produced by individual speakers are not directly related to the organization of the cognitive machinery that produces the language behaviors. (The behavior of a computer that is turned off every evening will have limitations due to the periodic stoppage, but such limitations ought not be confused with limitations intrinsic to the structure and manner of operation of the computer itself. Similarly, the fact that people die or tire or get bored with the sound of their own voices does not directly reflect on the nature of those parts of their minds that are devoted to the production of speech.) We can look at the whole argument from another point of view. The claim that there is an infinite set of English sentences is not really an assumption, it is rather a theorem derived from more basic assumptions. One of the most important of these assumptions is the claim that English sentences are generated by recursive application of certain syntactic operations. The assumed recursive nature of the generator is essential to the argument that the set of sentences is large without limit. Thus, recursivity is a more fundamental presupposition with respect to the claim of creativity than is the infinitude of possible sentences.

We do not become convinced that English sentences are creative behavior by being given infinitely many sentences or by being shown an infinitely long sentence, but by becoming convinced that, given any English sentence, certain operations can be applied which result in a string which is also an English sentence different from and longer than the original. It is the fact that language requires a theory accounting for this degree of complexity, rather than the unlimitedness of the set of English sentences, which is the basis of the claim that language is creative behavior.

Recursive mechanisms are, of course, required to account for most of the kinds of behavior Wickelgren wants to call noncreative ("animal 'speech' and mimicry, and all elementary skilled motor movements such as throwing a ball, running, walking, jumping, turning one's head, moving one's hand from one location to another, etc. [p. 2]"). No one would try modeling the free-hand drawing of a spiral on a piece of paper without recursively applicable operators like trigonomic functions. Such operators are essential here, as they are in the language case, even if one merely wants to give a description of the behavior itself, let alone the mechanisms that produce it. Behaviors which are complexly structured to an extent which forces us to consider them as creative are exceedingly common—indeed we do not know of *any* behaviors which are clearly noncreative, excepting possibly, spinal reflexes.

Although the preceding sentence may be an overstatement, it appears incontestable that Wickelgren's class of noncreative behaviors must be smaller than he suggests. If we consider for a moment the behavior of drawing curves on a piece of paper, as suggested above, we can see that the set of such curves is not only large without bound, but has (with the acceptance of a few assumptions similar to those necessary to prove that there is a potential infinity of grammatical English sentences) cardinality greater than Aleph Null. This is not merely a problem in nondenumerability, and it cannot be dismissed by alluding to the hypothetical "grain" of resolution in each sensory and motor dimension which is supposed to limit the finite set of classes at the noncreative level in Wickelgren's account. There are surely potential infinitudes of skilled motor acts in the same sense that there are infinite curves and infinite sentences. Thus, even if we accept only the criteria we presume Wickelgren used for creative behaviors, almost all of what we call behavior is sure to be included in that category.

As we see it, Wickelgren's distinction appears to be vague and difficult to apply.

Further, it seems to isolate only a tiny portion of behavior. It may be true that association theories are useful within that small domain, but it is clear that the classes of behavior we are discussing are not included.

MODELS FOR SPEECH PRODUCTION

Wickelgren gave four models for speech production and examined them in relation to the size of the central nervous system and the fact that speakers correctly order phonemes in words. He also claimed to deal with the facts of coarticulation phenomena. We find the context-sensitive associative model that he favors to be unsatisfactory and the nonassociative model that he pits against it remarkably weak. We will consider these briefly.

Context-Sensitive Associative Model

The problem in an associative model of speech is to cope with the stubborn fact that each phoneme cannot be used to call out the next one in typical associative chain fashion because all the phonemes have to be available in a huge variety of combinations. Wickelgren's strategy is to avoid this problem by creating enough versions of the phonemes to make them all distinctive in their various contexts. This is no mean feat. It entails increasing the number of speech production entities from about 50 to about 1,000,000. He regards this as acceptable since the nervous system has more than enough neurons to accomplish this. It is even hinted that this move is meritorious because it seems probable "that the nervous system generally uses a very large number of elements as a basic feature of its approach to control systems problems [p. 7]."

In essence, a new speech production entity is created for each speech sound in each context to capture its dependency on the speech units which precede and follow it.[3] If, however, coarticulational effects are found as near as two phonemes away from the speech unit of interest, the system becomes absolutely impossible even under the weak constraints used by Wickelgren; such a circumstance would require as many unique speech sound representatives as there are neurons in the central nervous system (100 × 100 × 100 × 100 × 100 = 10,000,000,000 separate identifiable units). Unfortunately for the theory, evidence suggests that such remote dependencies exist. For example, Kozhevnikov and Chistovich (1965, pp. 124 ff.) report that the initial /s/ in /stru/ already shows the influence of the choice of the vowel which follows *three* phonemes later. It seems likely to us that the more we perform careful finegrained analysis of *any* series of behavior sequences, the more rapidly we will push to disastrous limits this strategy of creating new units to do the associative job. This is as true at the level of words and lower levels as it is at the level of syntax.

The situation may be worse than the preceding paragraph may suggest. Wickelgren's numerical gymnastics are based on an assumption that he knows is tenuous, the assumption that one word or one elementary motor response equals one neuron. Lashley and his successors demonstrated that the cortex is not neatly cut up into tiny pieces of tissue with specific functions assigned. Cortical lesions associated with aphasia do not result in the loss of specific words, they result in the loss of language functions generally (Schuell, Jenkins, & Jiménez-Pabón, 1964). The concept of mass action may be only partially correct, but to suppose that it is entirely wrong is to ignore the body of our understanding of cortical functioning.

By the above, we do not mean to imply merely that each word requires more than one neuron for its storage. If current RNA coding models of memory are correct, it is perfectly conceivable that thousands of bits of information might be stored in a single neuron. Alternatively, if Hebb's suggestions regarding cell assemblies and phase sequences are correct, it is patterning of activity rather than specific loci that is important. The point here is that the notion of associative storage in specific locations and the retrieval of "units" is a model that we know to be impossible. Thus, it cannot be of any help to us either constructively or destructively to consider the number of neurons available in the central nervous system, since such a tally cannot have any meaningful relation to any plausible theory.

---

[3] One hundred preceeding sounds × 100 units × 100 succeeding sounds = 1,000,000 unique units. There are about a hundred basic units because Wickelgren is forced to treat vowels with different stress essentially as different phonemes. We will not stop to contest this strange tactic since our point here is that the whole argument is vacuous whatever the number of elementary units.

At this point, we had hoped to give a detailed account of the failings of Wickelgren's model in terms of well-known facts about linguistic structure and speech production, but it rapidly became clear that, on the one hand, these failings are so pervasive and, on the other hand, the theory is so inexplicit that it would be impossible to discuss the limitations of each of the possible versions of the theory in a paper of this size. The alternative we have adopted is to attempt a general characterization of the empirical limitations of Wickelgren's model with a few important examples.

The empirical inadequacies of Wickelgren's model are of two kinds, which may be characterized as (*a*) inability to account for the systematic relationships between events at the phonetic level and events at higher linguistic levels and (*b*) inability to account for the systematic interactions among events within the phonetic level. While these failings are abstractly equivalent, we will deal with them separately, in turn.

The context-sensitive associative memory model is unable to account for orderly syntactic relations among phonological events simply because this model, being a first-order Markov generator, allows of no syntax, nor, of course, of any influence of syntax on elementary motor responses.[4] Thus, for example, there is no way to account for the orderly variation of the English plural. The fact that the final phoneme of the noun stem determines whether the plural will be realized as /s/, /z/, or /iz/ cannot be represented in this theory. (We feel that it should be clear to the reader why this is so, and why this is a critical failure for the theory. Clearly Wickelgren's theory does not have laws against regularities, it simply ignores them and along with them most of the known regularities of its domain; regularities such as those underlying the ability of children to acquire and employ a variety of "pig Latins"; the ability to imitate an "accent," a dialect, or a lisp; the ability to modify articulatory behavior in order to talk while performing other acts with the mouth such as eating, chewing gum, holding a pencil in the mouth, or smoking; the ability to whisper, to shout, to speak rapidly or slowly; the ability of the ventriloquist to speak a modified version of the speech code; and so on. We take it as obvious that a theory does not succeed by being immune to evidence but by being vulnerable to evidence and surviving.)

Further, we must take cognizance of the fact that the failure of the theory to capture phonological regularities means that such regularities cannot be used to reduce the load on memory in human performance. Wickelgren's English cannot contain the abstract morpheme "plural" as part of a direction to a phonological unit since there is no way to predict what the plural of "cat" will be; we must look it up, just as we now must look up the plural of "mouse" on a list of exceptions.

To imagine using such a language in which there are no regularities and everything is an "exception" is staggering. It is not sufficient to remember the noun and apply the rule to get the plural—one must remember both the singular and the plural of *each* noun. This is also true for the past tense (judge–judged: /dʒʌdʒ/–/dʒʌdʒd/; slip–slipped: /slɪp/–/slɪpt/) and for the present progressive (run–runs: /rʌn/–/rʌnz/; hop–hops: /hap/–/haps/) in which the regularities are also conditional on voicing. Likewise, the regularities of English stress assignment (as in the stress relationships among various noun stems, e.g., telegraphic, telegraphy, telegraph) would disappear, at least for psychologists.

This inability to deal meaningfully with the elementary facts of English morphology seems to follow reasonably from a characterization which holds phonology and syntax to be so completely independent that it becomes possible to suppose that phonology can be noncreative. This independence of levels is an intersection of two unfortunate positions—the claim of American Structural Linguists that syntax and phonology are independent, and the tendency of American Associationists toward theories in which various levels of a hierarchical structure are presumed to operate independently. Combined in Wickelgren's theory, they give rise to a model which systematically ignores essentially everything that is known about words except the fact that the phonological events come in a particular order.

---

[4] For discussions of the inability of such models to cope with syntactic regularities, see Chomsky (1957, 1963).

The model also systematically ignores everything known about the influences of phonological events on other phonological events, Wickelgren's comments about coarticulation notwithstanding. For although he claims that one evidence of the superiority of the context-sensitive associative memory model is its ability to "account for" context effects, he proceeds to account for merely the fact that context effects occur, with no concern at all for the regularities of the *way* in which they occur. In Wickelgren's model, the structure of each context-sensitive phoneme has no relationship to the structure of any other context-sensitive phoneme, because context-sensitive phonemes *have* no intrinsic structure. Wickelgren's model cannot account for the fact that voiceless stops /p/, /t/, /k/ lose their aspiration when pronounced after initial /s/, because their is no room in the model for abstract phonetic categories like "voiceless," "stop," "aspiration," etc. Coarticulation is not described as a set of regularities like "when voiced stops follow a vowel, the vowel is lengthened," but rather as a list of unrelated articulating patterns, one for each context-sensitive phoneme.

Wickelgren hopes somehow to bring the phoneme back into the picture since he supposes that the "context free phoneme" is to have some role in learning how to speak, but he holds that such phoneme representatives need not play a role in control of articulation; nor are they employed in speech recognition. Thus, it seems to us that he sacrifices any hope of representing important regularities. He does not tell us in any clear way how phonemes are to be introduced in the theory or in behavior or how they can be associatively learned, aside from an appeal to the notion that the phoneme is an allophone class based on similarity of articulation (presumably sensed in the nervous system, but in ways which are not even mentioned).

In brief, it seems to us that theory is being developed in the wrong direction. In place of depth of theoretical construction, it substitutes massive multiplication of allegedly basic entities. Not only does this strategy violate our intuitions about what is "elegant," "direct," and "plausible" (all Wickelgren's terms), it also makes it difficult to describe known systematic effects in speech production. This kind of theory would presumably lead investigators away from the discovery of further systematic effects by removing the theoretical reasons for hunting such phenomena. It is this very general and potentially dangerous aspect of the theory to which we object most strongly. It should be clear to the reader who is familiar with Wickelgren's paper that the above comments apply as well in substance to the two other associative theories that he considers.

Nonassociative Theory

The nonassociative theory suggested by Wickelgren is unbelievably weak. Essentially, he postulates a serial memory which is more or less directly connected to a simply conceived motor control system. Serial order is made possible by explicit motor commands which are stored in just the order they are to appear in the motor output. The "nonassociative memory" is scanned serially and the appropriate pieces of behavior are emitted serially.

No "cognitive" psychologist would propose a model of this sort or would entertain one such as this for more than a few moments. The fundamental difficulty with this model, as with *all* of the models that Wickelgren considers, is that it is a *memory* model. (By "memory model" we mean any model in which the structure of behavior comes about by virtue of the way in which the relevant motor command sequences are recalled from a hypothetical memory.) There is no reason to believe that *any* memory model is capable of doing the job.

We feel that underlying this point is a very broad issue that often becomes obscured in the debates about details of models. That issue concerns the general acceptance of association theory which, as Lashley pointed out, provided the foundation for such disparate paradigms as structuralism and behaviorism and almost completely dominates American psychology today. Psychologists fail to consider productive models of behavior because they have accepted the assumption of empiricism and associationism that all, or nearly all, of the complex behaviors are learned behaviors. We tend to believe that the environment has provided us with exemplars and experiences from which we have derived response repertoires keyed to appropriate stimuli, which we now carry about with us in our memories. On this view, there are three main processes of interest: learning (acquisition), retention (storage), and recall (retrieval). Since we assume that speech is learned, the appropriate questions become: In what form have we stored the speech elements? and How do we gain access to them for retrieval? (In a more general sense, an associative theory of a given type of behavior is essentially a theory of the part of memory which is devoted to that type of behavior.) Thus, Wickelgren can exhaust the set of obvious models for speech production (and many other serially ordered behaviors) by considering four *memory models*.

The alternative that we support is that responses are not "looked up" in a vast rote memory and emitted serially but, rather, are generated by a complex processor operating on abstract entities and emitted by a very complicated series of transformations—transformations that by their very nature cannot even be considered in an associative theory.[5] We believe that Wickelgren's nonassociative memory model will not work for the same reason that his other models will fail—because they are memory models that attempt to capture an essentially creative behavior with a necessarily uncreative (in the technical sense) theory.

Wickelgren is apparently ready to conceed (for the moment at least) that syntax is creative and nonassociative. His failure to extend the same argument to speech production is attributable, perhaps, to a misunderstanding of the nature of productive theories. He says it would be "absurd to claim . . . that the full apparatus of a generative grammar is necessary for serial order in noncreative behavior such as the pronunciation of a single word [p. 2]."[6] It *would be* absurd, but not for the reason Wickelgren has in mind. It would be absurd because fundamentally the formal nature of a generative grammar is in essence not the formal nature of a psychological theory. No grammar could be a psychological theory, let alone a correct one. A grammar is a theory of the structure of behavior. A psychological theory is a theory about the nature and functioning of the cognitive "machinery" which produces behavior. But the generative grammar indicates the magnitude of complexity that a correct model of the behavior will have to embrace.

How complicated is the production of a single word? At the phonemic level, the word can be represented as a matter of some small amount of information; but at the motor command level, it is represented by a much larger amount. It is reasonable to suppose that the brain takes advantage of the compact and efficient coding offered by some level approximating that of the phoneme. And it also seems reasonable to us to suppose that the brain does a good deal of intricate and productive processing in converting this simple phonemic representation into the vastly more complex motor commands. It is not at all absurd to suppose that an adequate theoretical description of this processing will be at least as intricate as a generative grammar. In this, we merely echo Mattingly and Liberman (1969) who considered this question and concluded:

---

[5] Perhaps at this point we should stress that we do not wish to exclude from consideration all models which involve memorial processes. It is possible to conceive of memories that are relatively creative, and such memories might play a large role in the production of behavior.

[6] By "generative grammar" we assume that Wickelgren means a generative grammar of the transformational (or some other relatively powerful) type, since it would seem to be relatively easy to show that this context-sensitive associative memory is at the same level of formal power as a right linear (generative) grammar. (His prefix "context sensitive" is unfortunately very misleading in this regard.)

> From the formal point of view, then, there is good reason to regard the acoustic signal as another linguistic level, and the conversion from the phonetic message to the acoustic signal as a process comparable, in an important sense, to the conversions at higher linguistic levels [p. 107].

For nonassociative productive models, the problem of serial order is not a serious one. There are myriads of possible mechanisms with sufficient power to produce sequentially ordered outputs. The problem that confronts the cognitive psychologist is, rather, to select from among the many possible mechanisms one which is a reasonable model of the human cognitive apparatus. It would be pretentious for us to offer any kind of specification here. What we urge is that interested psychologists devote their attention to studying the production and perception of speech with the objective of discovering the rules that govern the phenomena and inventing the theoretical apparatus needed to make substantive advances rather than attempting to build models of a very restricted and generatively weak class simply because they are sanctified by tradition.

CONCLUSION

Lashley (1951) knew that the nervous system did not operate on the basis of one-neuron-to-one-function and that simple associationism was unsupported by what he knew of the brain's activity. He illustrated his position behaviorally by presenting sets of serially organized behavior that elementary associative theories could not explain. Wickelgren (1969) has attempted a rejoinder to Lashley by postulating a class of noncreative behaviors, of which speech production, at the level of the phoneme, is an instance. He proposes a context-sensitive associative model to account for serial ordering in behaviors of this class.

The present paper objects to the distinction between creative and noncreative behavior on the grounds that Wickelgren's distinction seriously misinterprets the creativity notion. Beyond that point we have objected to theories of speech production based on the associative tradition, pointing out that Wickelgren's model has certain serious disadvantages and that memory models in general (an inheritance from the Associationists) will prove insufficient. It was urged that we let the phenomena of behavior and neurology restrict our classes of models rather than allowing them to be dominated by tradition.

REFERENCES

CHOMSKY, N. *Syntactic structures.* The Hague: Mouton, 1957.

CHOMSKY, N. Formal properties of grammar. In R. D. Luce, R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology.* Vol. 2. New York: Wiley, 1963.

CHOMSKY, N. *Current issues in linguistic theory.* The Hague: Mouton, 1964.

CHOMSKY, N. *Cartesian linguistics.* New York: Harper & Row, 1966.

CHOMSKY, N., & MILLER, G. A. Introduction to the formal analysis of natural language. In R. D. Luce, R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology.* Vol. 2. New York: Wiley, 1963.

KOZHEVNIKOV, V. A., & CHISTOVICH, L. A. *Speech: Articulation and perception.* Washington, D. C.: Joint Publications Research Service, 1965.

LASHLEY, K. S. The problem of serial order in behavior. In L. A. Jeffress (Ed.), *Cerebral mechanisms in behavior.* New York: Wiley, 1951.

MATTINGLY, I. G., & LIBERMAN, A. M. The speech code and the psychology of language. In K. N. Liebovic (Ed.), *Information processing in the nervous system.* New York: Springer-Verlag, 1969.

SCHUELL, H., JENKINS, J. J., & JIMÉNEZ-PABÓN, E. *The problem of aphasia in adults: Diagnosis, prognosis and treatment.* New York: Hoeber, 1964.

WICKELGREN, W. A. Context-sensitive coding, associative memory, and serial order in (speech) behavior. *Psychological Review,* 1969, **76**, 1–15.

(Received June 25, 1970)

Psychological Review
1971, Vol. 78, No. 2, 130–140

HYPOTHESIS THEORY AND NONLEARNING DESPITE IDEAL S–R-REINFORCEMENT CONTINGENCIES[1]

MARVIN LEVINE[2]

State University of New York at Stony Brook

Hypothesis theory generates a few theorems about adult humans failing to reach criterion in discrimination-learning tasks. First, their performance should match behavior on tasks which are intrinsically insoluble but otherwise comparable to the normal tasks. Second, for certain stimulus sequences, performance on both the unsolved and insoluble problems should be below the chance level. The data confirmed both of these predictions. The theory also stipulates the conditions for producing nonlearning by motivated *S*s. According to the theory, as long as *S* is sampling from a set of incorrect hypotheses (i.e., a set lacking the solution) he will show no learning. No response strengthening will occur in this circumstance even if the solution consists in the simplest contingencies (e.g., *E* says "right" when *S* says "A," and "wrong" when *S* says "B"). A series of experiments validated the prediction. The relation of these results to random-feedback effects, the awareness controversy, and the Einstellung phenomenon is discussed.

A college student appears at the laboratory and is given a simple learning task. He is instructed to respond to each of a series of stimuli with any one of a limited set of responses. After some (preselected) response the experimenter (*E*) says "right," after the others he says "wrong." The well-known experimental result is that the selected response increases in relative frequency.[3] For more than half a century this finding has been so ubiquitous that it has been twice codified, as the Law of Effect and as the Reinforcement Principle.

The phenomenon, however, is not absolutely reliable. There are at least two classes of results indicating that learning does not always take place. The first comes from studies concerned with the role of awareness in learning. Several authors (e.g., Dulaney, 1961; Spielberger & DeNike, 1966) have identified subsets of subjects (*S*s) showing no change in response probability throughout the task. The second comes from investigations of transfer of learning from a first to a second discrimination problem (e.g., the reversal-shift procedure). The authors typically report the elimination of several *S*s who, after an inordinately long number of trials, still had not learned the first problem. It appears, then, that nonlearning certainly occurs. The experiments below will demonstrate that it occurs even with well-motivated *S*s for whom all the relevant information is available and unambiguous. The purpose of this article, however, is not merely to substantiate this finding, but to present a theory which accounts for this and related results.

Hypothesis (*H*) theory has been vigorously developed in recent years (for reviews see Levine 1969; Trabasso & Bower, 1968).

---

[1] The experiments described here and the preparation of this article were supported by Research Grant MH 11857 from the National Institute of Mental Health. I wish to express my appreciation to William Glassman who performed the first experiment and to O. Ernest Andrews who supervised the last set of experiments.

[2] Requests for reprints should be sent to Marvin Levine, Department of Psychology, State University of New York at Stony Brook, Stony Brook, New York 11790.

[3] More typically, *E* says "right" only to certain responses in the presence of certain stimuli. For example, *E* may say "right" to *S*'s response "alpha" only if triangles are part of the stimulus complex. Otherwise he says "right" only to the response "beta." While it is always more accurate to speak of *E*'s feedback as contingent on certain responses *in the presence of certain stimuli,* the italicized phrase will generally be omitted. It should, however, be understood as given throughout.

The work suggests that in discrimination- or concept-learning laboratory tasks, adult human behavior is the result of H testing. Novel analyses (backward learning curves; blank-trial probes) imply that S has a set of Hs from which he repeatedly samples until he selects the correct H. Some detailed specifications of the theory will be given below, along with a review of the data on which the specifications are based. However, it will be useful first to have a clear description of the initial task to which the theory will be applied.

Several 12-trial multidimensional simultaneous-discrimination problems were constructed. Each trial was initiated by the display of two stimulus patterns (cf. Figure 1). One pattern consisted of a set of values from each of eight two-valued dimensions; the other pattern consisted of the complementary values of the dimensions. From trial to trial the dimension values shifted from one pattern to the other so that on some trial the T might be large, surrounded by one solid circle, etc. The shifting of values from trial to trial followed these rules: (*a*) on each pair of adjacent trials, values from four dimensions remained paired together (i.e., were on the same side for both trials) while the values from the other four dimensions were changed (shifted sides from one trial to the next); (*b*) no two dimension values stayed paired with each other for more than three consecutive trials (e.g., A's could not be large more than three times in a row, circles could not be paired with the solid underline more than three times in a row, etc.); (*c*) for any dimension value there was one and only one other value paired for the three trials.

The experimental problems were basically of two types: 12-trial 8-dimensional (8-D) problems with legitimate solutions and 12-trial 7-D insoluble problems. The latter were formed by selecting the solution from 8-D stimuli and by then eliminating the relevant dimension. Suppose, for example, that (before the experiment began) "single spot" was selected to be the solution. The E noted on which side the single spot occurred on each of the 12 trials—choice of this side would be correct. He then made 12 new stimuli, identical to the old except that the spots were omitted from both of the stimulus patterns. The S received eight such insoluble problems distributed throughout 24 legitimate problems.

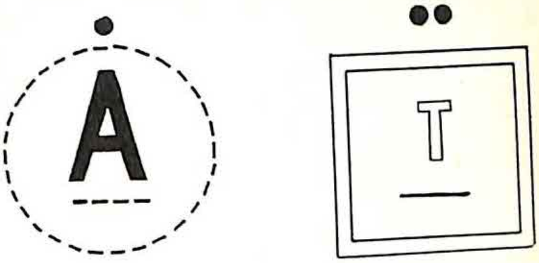


FIG. 1. A simultaneous-discrimination stimulus constructed from eight dimensions (letter, A–T; letter size, large–small; letter color, black–white; underline, solid–dashed; border shape, circle–square; border number, one–two; border texture, solid–dashed; spots, one–two).

The Ss were first shown examples of the stimuli and were carefully instructed about the eight dimensions. They were told that they were to select one of the two patterns and that after each choice E would indicate the correctness of that response. They were further told that the solution (i.e., the basis for responding correctly) on any one problem was one of the 16 attribute values (8 dimensions × 2 values per dimensions) and that they were to try to be correct as often as possible. The S was then given several practice problems (both 7-D and 8-D—all soluble) followed by the 32 (24 soluble, 8 insoluble) experimental problems.[4]

In the results, three types of problems can be characterized: legitimate 8-D problems which are solved (defined as having correct responses on Trials 10, 11, and 12)—these will hereafter be ignored; legitimate 8-D problems which are unsolved (defined as having an error on one of Trials 10, 11, or 12); and insoluble 7-D problems, which, of course, S cannot solve. It is the comparison of performance on these last two types of problems which is of interest. Before presenting this comparison, however, the theory will first be reviewed.

---

[4] For a more complete review of this experiment, see Glassman (1970).

THE THEORY OF INSOLUBLE AND UNSOLVED PROBLEMS

The necessary assumptions are:

1. The basic assumption: The *S* selects an *H* from some universe and responds on the basis of that *H*. For example, *S* may predict that circles are the basis for choosing. He would then choose the side containing the circle. This is a primitive axiom, although support for the second part, that the *H* directs the response, is found by inserting blocks of blank (i.e., no feedback) trials into normal problems. Frankel, Levine, and Karpf (1970) have shown that *S*s virtually always respond to a single attribute value for as many as 30 consecutive blank trials.

2. The composition assumption: The universe from which *S* samples contains only those *H*s corresponding to the stimulus levels. The *S* predicts that "large" or "T," etc., is correct. He does not employ *H*s about conjunctions of cues or about complex sequences. This assumption, clearly, is a function of *E*'s instructions and of the preliminary practice problems. By altering these, the assumption can be changed without affecting the rest of the theory. The last set of experiments described below employs a different composition assumption (and, concomitantly, entails different pretraining). The important feature of the present assumption is that for insoluble 7-D problems, *S*'s universe consists of 14 *H*s.

3. Presolution responding: Prior to solving the problem (as manifested by a string of successive correct responses) *S* samples neither the correct *H* nor the other *H* on the same dimension. Thus, if "large letter" is the solution, before the criterion run—or during unsolved problems—*S* tries neither the *H* "large letter" nor "small letter." This assumption is based on fairly direct experimental support. Levine, Miller, and Steinmeyer (1967), probing for *S*'s *H* after each trial, showed that prior to solving, *H*s from the relevant dimension virtually never appeared. The important feature of this assumption is that during unsolved 8-D problems, *S*'s universe consists of 14 *H*s.

It is hoped that the reader has noticed the congruity between the last two assumptions: During both insoluble 7-D and unsolved 8-D problems, the *S* is sampling from a universe of 14 wrong (irrelevant) *H*s. In both cases, the correct *H* (H^+) and the complement of H^+ are outside the set from which *S* is sampling. This similarity in theoretical state in turn implies that the behavior under the two conditions should be similar. This, then, is the first prediction of the model, that the trial-by-trial behavior should be the same during both an insoluble 7-D problem and an unsolved 8-D problem.

The next two assumptions will permit further refinement of this prediction and some specification of the details of that trial-by-trial behavior.

4. Affirmative feedback: If *S*'s response is followed by a positive outcome (e.g., *E* says "correct"), he keeps his *H* for the next trial. Probing for the *H* after each trial, Levine (1969) showed that, with the present procedure, this assumption almost always held ($p \approx .95$).

5. Negative feedback: If *S*'s response is followed by a negative outcome (e.g., *E* says "wrong"), he abandons his last *H* and resamples. Levine (1969) showed that this assumption also virtually always held ($p \approx .99$). Several proposals have been made concerning *S*'s mode of resampling following a "wrong." A few of these will be described here. All of these alternatives can be ordered along a memory continuum describing how much *S* remembers from past trials. Each proposal will be illustrated by an *S* sampling from a universe of 14 wrong *H*s.

(*a*) Reject 0: At one extreme, *S* might remember nothing, not even the *H* just disconfirmed. He would, in effect, return this *H* to the set and would resample from the full universe of 14 *H*s. This assumption, first made by Restle (1962), has been shown to be generally false (Levine, 1966; Restle & Emmerich, 1966). It is included here, however, as one end of the memory continuum.

(*b*) Reject 7: The *S* might remember all the *H*s corresponding to the stimulus pattern he just chose and might recognize that they are all disqualified by the negative feedback. He resamples, in effect, from the seven *H*s consistent with the information on the trial just past. Also, he does not utilize information from earlier trials. Gregg and Simon (1967) dubbed this the "local consistency" rule. Trabasso and Bower (1968) have employed this assumption.

(*c*) Reject 11: The *S* might sample from *H*s consistent not only with the just-past trial but with the preceding trial as well. With the stimulus sequences of the present experiment, this always means that *S* ignores 11 *H*s when resampling after a "wrong." This two-trial consistency check is part of an assumption proposed by Trabasso and Bower (1966).

(*d*) Reject all: The *S* might remember more than just two previous trials and could, therefore, eliminate more incorrect *H*s. At the extreme, his memory would be perfect and he would always sample among logically correct *H*s. This corresponds to the performance of focusers as described by Bruner, Goodnow, and Austin (1956, pp. 129–134). This poses an interesting psychological problem in the present context. The stimulus sequences are such that by the fourth trial of a problem all 14 *H*s are disconfirmed. The focuser facing an insoluble problem has an empty universe after Trial 4. Some suggestions concerning the behavior of *S* in this circumstance will be made later.

This is all the theory necessary to derive additional properties of *S*'s behavior during insoluble and unsolved problems. The prediction was already stated that during both types of problems, the protocols should match each other. Further predictions come from applying the theory to the sequence of stimuli employed. Within a problem, the stimuli, it will be recalled, were selected not completely randomly, but to fit stringent counterbalancing requirements. Because this counterbalancing held for every problem, it is possible to derive a predicted performance curve when the response rules described in the theory are applied to these stimulus sequences. In general, the predicted performance curve depends on the memory assumption held for a resampling *S* (i.e., upon variants *a*–*d* of Assumption 5). The family of memory assumptions generates a corresponding family of predicted curves. Figure 2 shows the curves predicted using each of the four variants.[5] The most striking feature of the predictions is that for all the variants of the memory assumption, the theory predicts that during both insoluble and unsolved problems, performance will be below chance. The theory, then, predicts that not only will the curves from both types of problems be the same, but that for both, fewer than half of the responses will be correct. Furthermore, correct responding will decrease during the first four trials and will then show an upswing. Again, these predictions hold regardless of *S*'s memory, that is, they hold for all the curves in Figure 2. The details of performance following Trial 4 is currently a matter for some speculation and will be discussed later.

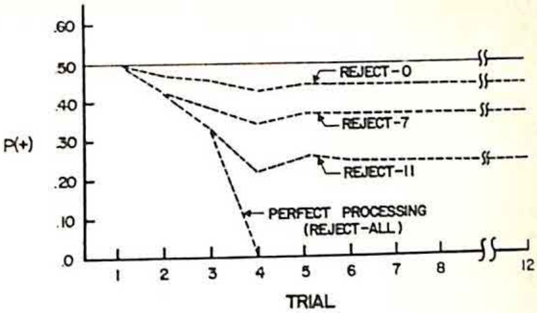


FIG. 2. The family of theoretical curves, each of which corresponds to the particular memory assumption added to the theory.

The data comparing 12-trial insoluble and unsolved problems are shown in Figure 3. It is seen there that the predictions are verified precisely. The curves show below-chance responding and the initial decrease. Most importantly, the curves are closely matched: At no trial do the pairs of points differ at the .01 level; at only one trial do they differ at the .05 level. This similarity holds despite the fact that the .05 confidence interval (indicated by the shaded band in Figure 3) is fairly narrow. The conclusion is compelling, therefore, that the relevant dimension is not being sampled by an *S* when he is faced with a soluble problem which he is not solving.

---

[5] A representative derivation of this type of theoretical curve may be found in Levine, Yoder, Kleinberg, and Rosenberg (1968).

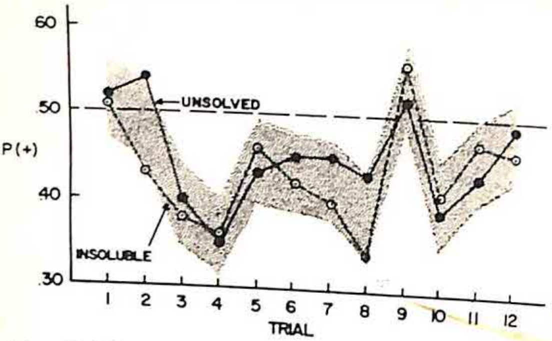


FIG. 3. The trial-by-trial proportion correct during unsolved and insoluble problems (the shaded band shows the .05 confidence interval around the pair of points at each trial).

THE EMPTY SET

The *S* who does not solve a soluble problem was described above with a general conception: He samples from a finite set of *H*s (14, in the particular case considered), and the correct *H* is outside this set. A basic question for the learning process concerns how *S* might eventually abandon this incorrect set and turn, at some later point, to sample the correct *H*. This question is intertwined with a question raised earlier. It was noted above that if *S*'s memory for disqualified *H*s is good enough, his set, which contains only incorrect *H*s, will soon become empty. What alternatives are open to an *S* when this happens, when the last remaining *H* proves to be wrong? Consider a homey example—a person is looking for his car keys. He thinks, "Perhaps I left them in one of my jacket pockets." He opens his closet and checks the pockets of the jackets hanging there (the pockets constitute his *H* set). Suppose a rapid check of all the pockets reveals no clinking sound of keys (the set goes to zero). One possibility is that he thinks "Did I skip some pockets? Perhaps I should have checked each pocket more carefully." That is, one possibility is that he goes through the set again. A hint that this is happening in the experiment just reported (see Figure 3) is in the cyclic pattern during the insoluble problems. A perfect processor starts with the full set of *H*s at the first trial and reduces the set until, by the third trial, he has eliminated everything but the one *H* correlated with the (absent) correct *H* for the first three trials. By the nature of the stimulus constraints, the probability that this *H* will lead to a correct choice on Trial 4 is zero. The perfect processor selects this *H* after Trial 3, is told "wrong" at Trial 4, and then has an empty set. If *S* checks again through the set at this point, one should see a rise in the proportion correct at Trial 5 followed by another minimum a few trials later. The cyclic character of the curves in Figure 3 suggests that this resurvey of the set might be occasionally occurring.

Another hint that *S*s review the *H* set after it becomes empty comes from an experiment by Erickson (1968), who used four-dimensional stimuli. Erickson presented blocks of three blank trials before each feedback trial as a technique for probing for *S*'s *H*. He disconfirmed the first *H* by indicating that the response to the first feedback trial was wrong. Both Erickson (1968) and Levine (1966, 1969) (also cf. Assumption 5, above) showed that *S*s do not immediately return this *H* to the set, where it may be resampled. The *H*, then, is out of the set. The twist in Erickson's experiment was that this *H*, for all remaining trials, was made the correct *H*. Thus, an *S* might select the large stimulus for the first four trials. After the fourth trial (i.e., the first feedback trial) the *E* says "wrong." Choice of the large stimulus is then the correct response for all remaining trials.[6] Conditions of the experiment facilitated information processing: not only was there a small number of dimensions, but the stimulus stayed on when feedback was presented, and there was a long intertrial interval (cf. Bourne, Guy, Dodd, & Justesen, 1965, for the effectiveness of these variables). In fact, in a standard control condition, in which feedback was consistent throughout the whole problem, the mean number of feedback trials to criterion was 4.1. In the experimental condition (with the first-trial reversal), 48 of 50 *S*s solved the problem. They required on the average only 2.3 feedback trials more than the control condition, approximately the result one might expect if *S*s typically reviewed the *H* set a second time. One reaction, then, when a set goes to zero is to go through it again.

---

[6] This is a loose parallel to Erickson's procedure. He, in fact, used a successive discrimination procedure. The description correctly conveys the relevant features, however, without laboring over the irrelevant differences.

For another reaction let us return to our friend bereft of his car keys. Suppose that he again goes through his jacket pockets for the keys, this time more systematically (i.e., in such a way as to retain in memory that every *H* was tried and rejected). The set is again empty. What other alternative is open to him? He can leave this set of *H*s and look for a new one. He can do this by scanning his memory ("What was I doing the last time I saw those keys?") or the present situation ("Could they be among that clutter of things on my desk?"). That new set may be fairly large ("Maybe they're in one of my pants pockets") or be very small consisting of virtually only the correct *H* (Of course! I was wearing my raincoat yesterday; they must be in my raincoat pocket"). A parallel within the experimental situation would occur if *S*, convinced that he is overlooking something, thinks back to *E*'s instructions, in an effort to retrieve some possibly forgotten dimensions.

Another reaction, then, when a set goes to zero is to abandon this set and to look for a new one. It is proposed that this is one important process underlying the shift from nonsolution to solution behavior. The description, while it is probably in keeping with everyday experience, is frankly speculative. Nevertheless, it is sufficiently important to the learning process that it warrants the status of an assumption, along with the five assumptions presented above.

6. Changing of sets (or subsets): The reduction of a set to zero *H*s serves as a signal to select a new set. As indicated above, this search for a new set may follow one or more reviews of the old set.

THE INFINITE SET

A mechanism was just provided permitting an *S* who starts a problem with the wrong *H* set to eventually solve the problem. That set will go to zero and *S*, perhaps after one or more reviews of the set, will look for a new one. Suppose the initial incorrect set can never go to zero? Suppose, that is, that *S* is sampling from a set which is infinitely large yet which does not contain the correct *H*? According to the theory above, *S* will never abandon that incorrect set. Only Assumption 6 allows for a change of *H* set, but it requires that the incorrect set become empty. An *S*, therefore, sampling from an incorrect but infinite set, will never solve the problem no matter how easy it is, that is, no matter how obvious the S–R and feedback contingencies are.

This implication was first explored by Ress (1965; for a summary see Ress & Levine, 1966). To realize the simplicity of the task employed, consider first the control group: Thirty *S*s each received a two-dimensional simultaneous discrimination. They went through a deck containing 115 cards, each card showing one large (2.5 centimeter) and one small (1.3 centimeter) circle. In typical fashion, for half of the cards, the large circle was (randomly) located on the right (or left) side. The *S* was instructed that he should touch the center of one of the two circles on each card, that *E* would say "right" or "wrong" after each response, and that he (*S*) was to try to be right as often as possible. The *E* said "right" whenever *S* touched the large circle, "wrong" whenever the small. In short, it was formally identical to the discrimination problem one gives to monkeys in the Wisconsin General Test Apparatus or to rats in the Lashley jumping stand. It is not surprising that these adult humans solved the problem in a mean of three trials.

The experimental *S*s received the same task. This was preceded, however, by a series of experiences intended to start the *S*s sampling from an infinite and incorrect *H* set. Sixty *S*s each received a series of preliminary problems using decks identical (except for shuffling) to the one for the main problem. The solution to these preliminary problems was always a position sequence. The sequences 2L-2R (two left-side responses followed by two right-side responses), 3L-2R, 1L-4R, 5L-1R, 2L-4R, and 3L-3R were presented.[7] The *E* announced the correct solution after each of these problems. It was assumed that this series of preliminary problems would fulfill the composition assumption that *S* was sampling from an infinite set containing *H*s dealing only with position sequences.

After the preliminary problems, *E* announced, as he had prior to each of the preliminaries, "Here is another problem" and presented *S* the reshuffled deck. This time, of course, the correct *H* was "touch 'large,'" an *H* not in the set of position sequences. The *S*s were run until they either gave 15 consecutive correct responses or made an error after Trial 100. The results were that these *S*s required a mean of 62 trials to solve this problem. Twenty-eight of the 60 *S*s did not solve it in 115 trials.

The *S*s, incidentally, were clearly motivated to solve these problems, and the feedback ("right" and "wrong") qualified as effective reinforcers. The preceding six sequence problems showed a typical learning-set effect. The mean trials to solution went from 65 trials on the first problem to 16 trials on the sixth.

At first glance it appears that the theory is strongly confirmed. The operations employed (i.e., the preliminary problems) for fulfilling the composition assumption that *S* is sampling from an infinite and incorrect set seemed clearly effective: The *S*s required an average of 62 trials to learn a discrimination any self-respecting monkey learns in a few trials. Corroborating testimony comes from postexperimental comments by *S*s. Virtually all acknowledged starting the final problem searching for sequence solutions. The results, however, do contain a puzzle for the theory. According to the theory, *S*s should not solve at all. If they are sampling from an infinite (and incorrect) set, from a set which does not become empty, then they should never select a new set. Yet more than half of the *S*s (32/60) solved the problem, that is, somehow selected a new set containing the correct *H*.

The puzzle can be resolved in one of two ways: (*a*) One may look for an additional assumption, postulating some other process by which an *H* outside the set can be sampled; or (*b*) one may question whether the composition assumption was fulfilled, whether the preliminary problems really did orient *S*s to sample from an *infinite* set of position sequences. A resolution to this puzzle, favoring the second alternative, was suggested in an analysis of sequence structures by Restle (1967). He noted that binary sequence cycles consisting of a single change (true of all the sequence solutions employed by Ress & Levine—cf. above) formed a particular subset of all sequences. The *S* could code these by what Restle called mandatory rules. More complex sequences, involving several changes in (binary) state within a cycle, required what he called optional rules. His analysis suggested that Ress and Levine trained the *S*s with only a subset of sequences, those having short cycles with a single shift within the cycle. This implies that at the final problem the *S*s might have held not an infinite set of all possible position sequences, but a finite set of similar and simple sequences. This set, of course, could go to zero. Some *S*s might then try a different incorrect *H* set (e.g., "Maybe it's a more complicated sequence") whereas others might come up with the set containing the correct *H* (e.g., "Maybe it has something to do with the circle").

The best hunch, then, was that Ress and Levine failed to obtain ideal results because the composition assumption had not been fulfilled. The obvious next step was to repeat the experiment employing more complex sequences. This, accordingly, was done. Two groups of 20 *S*s each received six preliminary sequence problems followed by the final test problem. The simple group received the same sequences as those employed by Ress and Levine; the complex group had sequences of varying complexity, the most complex sequence consisting of a cycle of 10 trials with five shifts during the cycle (e.g., LLRLLLRLRR). Except for this variation between the two groups, the task was identical to that employed by Ress and Levine.

---

[7] Two other variables were manipulated: The *S*s had either three or six preliminary problems; the sizes of the circles differed (2.5 vs. 1.3 centimeters and 2.5 vs. 2.0 centimeters). These variables had no significant effects. The data are combined here.

Thirteen of 20 *S*s in the simple group solved the final problem, thereby replicating the earlier results; in the complex group, however, only 3 of the 20 *S*s solved the problem. The difference is significant ($\chi^2 = 8.44$, $p < .01$). This experiment, then, suggests that the complex sequences produce more successful fulfillment of the composition assumption. Also, this dramatic demonstration of *S*s failing to learn the most obvious associations, further validates the *H* model described above as a description of the learning (and nonlearning) process.

Before considering the results of the complex group in greater detail, an additional experimental variation will be described. In the experiments above, *S* touched one of two circles. One of these circles (e.g., the larger) was the stimulus entering into the contingencies. One might propose an exceedingly simple reason why *S* failed to learn: he did not look at these circles. The *S* might, by this line of reasoning, touch the right or left side of the card without seeing the circles. To put it another way, Spence's (1945) criterion, that an observing response must occur before the data can be considered relevant to any learning theory, may not have been met.

Another experiment, therefore, was undertaken. The primary change from the preceding experiments was the replacement of the two circles on each card by two letters, A and B. Also, every *S* received the six complex sequences employed with the complex group, above. Two groups of 16 *S*s each were run. The touching group, as before, was told to touch one of the two letters; the speaking group was told to indicate each choice by saying one of the two letters. Note two features about the speaking group during the final problem: First, there is no way then they can avoid seeing the stimuli; second, the contingencies are even more simplified. Instead of discrimination learning, this is now the reinforcement paradigm in its simplest form: Whenever *S* says "A" *E* says "right"; whenever *S* says "B" *E* says "wrong."

There was essentially no difference between the two groups. Four *S*s in the touching group and three *S*s in the speaking group solved the problem. This result, along with the observation that *S*s in the touching group virtually always directed their gaze toward the letters, eliminates any concern that the nonlearning was an artifact of nonstimulation.

Three groups, then (the complex, touching, and speaking), were run in which the composition assumption appeared to be fulfilled. Of a total of 52 *S*s, 81% (42 out of 52) did not reach the criterion of 15 successive correct responses within 115 trials. It is possible, of course, that these 42 *S*s, though they failed this fairly extreme criterion, may nevertheless have shown some response strengthening. There is, however, no indication of anything but chance occurrence of the correct response during the 100 trials of the problem. On Trials 91–100, for example, 53% of the responses were correct. Despite a large *N* (42 *S*s × 10 responses per *S*), this is not significantly different from 50%. It appears clear that these *S*s had not switched to the set containing the correct *H*.

The remaining 10 *S*s did solve the problem. Why these *S*s might have been exceptional is suggested by comments made by two of them. One said that the whole experiment had reminded her of Luchins' water-jug experiment, which had recently been reviewed in her course; the other said that he had been looking for a trick—"psychologists always trick you." These comments imply that the preliminary problems did not always succeed in restricting *S*'s *H* set to position-sequence *H*s.

DISCUSSION

Two types of experiments have been presented demonstrating nonlearning. The first experiment showed that an *S* who had not learned behaved as though the relevant stimuli were literally absent. The second set of experiments, by reducing stimulus complexity to a minimum and by guaranteeing that *S* observed the relevant stimuli, demonstrated nonlearning in a more dramatic form. Nonlearning was shown also to be not a sporadic occurrence by an occasional *S*, but rather an outcome of definite problem-solving processes, susceptible to theoretical analysis and to manipulation. According to the analysis, nonlearning can be produced—despite a well-motivated *S* and consistent S–R feedback contingencies—by making *S* sample from an *H* set which does not contain the correct *H*.

It should be stressed that the theory employed was not ad hoc. A "theory of nonlearning" was not improvised to account for the data. Rather the data conformed to predictions derived from a general theory of hypothesis-testing behavior. The theory had already been employed to account for behavior during blank trials, for the discontinuity in the percentage of correct responses during simple learning, for cue redundancy and cue reversal, etc. The theory was here enlarged to include the concepts of empty and infinite sets, but this broadening followed plausible lines.

This expanded theory and the experimental results have relevance to three other research topics: the effects of random feedback on subsequent discrimination learning, the awareness controversy, and the Einstellung effect.

Random feedback: Levine (1962) demonstrated that the learning of very simple discriminations was clearly retarded if *E* said "right" or "wrong" randomly for the first few trials. This result was also shown to hold for concept learning by Mandler, Cowan, and Gold (1964) and for complex (6-D) discrimination learning by Holstein and Premack (1965). Levine suggested that some *S*s rejected the simple *H*s (including the solution) during the random feedback phase and were sampling from another subset (e.g., sequence *H*s) well into the discrimination phase. Holstein and Premack made a similar suggestion. Behind these suggestions was the tacit assumption that if *S*s were testing *H*s from an incorrect set, then they would not learn the correct contingencies. No attempt was made to demonstrate either that random feedback caused *S* to sample from an incorrect *H* set or that such a sampling would hinder the subsequent simple learning. The second type of experiment presented above, however, relates directly to the latter assumption. It demonstrated that by setting *S* to sample *H*s from an incorrect set, the learning of a simple contingency was indeed delayed. The operation of random feedback is well viewed, then, as a gross technique for manipulating the sets from which *S* samples. The resulting retardation in simple learning follows from *H* theory as outlined above.

The awareness controversy: In recent years several hundreds of *E*s have challenged or defended the reinforcement principle under the banners of the awareness controversy. Two extreme partisan positions in this controversy may be discerned. One (Dulaney, 1968; Spielberger & DeNike, 1966) is that the reinforcement principle is irrelevant, that only when *S* becomes "aware" of the contingencies will performance increase. The other (Dixon & Oakes, 1965; Verplank 1962) holds that the response strengthening occurs automatically, that awareness is an irrelevant or, at best, an independent process. Postman and Sassenrath (1961) took up a middle position, as have recently several other theorists (Bandura, 1969; Kanfer, 1968; Krasner, 1967). The view of Postman and Sassenrath is that both reinforcement and awareness contribute to the response-strengthening process, the former weakly but persistently, the latter strongly but rapidly, that is, in a discontinuous way. The results presented above are relevant not only to the extreme views but to this last moderate statement as well. These results demonstrated unequivocally that the reinforcement principle can be completely nullified. Establishing effective contingencies is not a sufficient condition for producing learning, that is, for increasing the probability of the contingent response.

The principle, so long a bastion of learning theory, may be defended in two ways. Silver, Saltz, and Modigliani (1970) suggest that learning via *H* testing and learning automatically via reinforcement may both occur but are mutually exclusive processes. According to this view, the procedures of the present experiments evoke the *H*-testing mechanism and thereby switch off automatic strengthening. These experiments, of course, cannot negate this possibility. Nor can any laboratory experiments which *S* might interpret as a problem-solving task. It may be, however, that the Silver et al. formulation will provide the ultimate resolution of the controversy.

The second defense of the principle comes through redefining the response. This is most easily seen in the second set of experiments. One may argue that because of the preliminary problems, saying "A" was no longer the functional response. Rather, the functional response, that is, that which was affected by the reinforcement, was *S*'s covert formulation of a position sequence. These covert formulations were so strengthened in the preliminary problems that they did not extinguish over the 100 trials of the test problem. Far from rebutting this defense; one need only note that it is a major conversion to *H* theory. Indeed, *H* theory may be regarded as the theory of these covert formulations. In the early stages of the development of *H* theory, Levine (1959, 1963) suggested that its most important feature was the redefinition of the response—"The *H*, rather than the specific choice response on a particular trial, is regarded as the dependent variable, i.e., as the unit of behavior affected by the reinforcements [Levine, 1963, p. 270]". Subsequently, *H* theory has had considerable refinement (e.g., the all-or-none character of *H* strengthening and weakening, the changes in strength of several *H*s simultaneously, the encoding mode—visual and verbal—of information, etc.), but the redefinition of the response remains the foundation on which the theory has been built.

Einstellung: The reader, like one *S* mentioned above, has undoubtedly noticed the similarity of the second set of experiments to the Einstellung experiment (cf. Luchins, 1942). In this experiment also, *S* has preliminary experiences which render him functionally blind to a subsequent simple solution. For example, Luchins presented *S*s with a series of paper-and-pencil maze problems. For the first nine problems, *S* was given a picture of a maze consisting of both a direct straight path leading from the start toward the goal box and other long circuitous paths. During these problems, the straight path always was blocked, that is, visibly ended as a cul-de-sac just before the goal box. One of the circuitous paths leading off from the right of the start box always eventually led to the goal. On the tenth problem, the direct path was connected to the goal. Nevertheless, the *S*s typically chose the round-about alley leading off from the right. In *H* terms, the analysis is clear: the preliminary problems led *S* to sample *H*s dealing with alleys off to the right. Until this set is exhausted, the *S* is insensitive to direct central alleys. This theory, incidentally, has an obvious implication not yet tested within the older Einstellung tradition. The theory implies that the larger the incorrect *H* set, the longer it takes that set to become empty, and the more durably is *S* functionally blind to the simple solution. This, in turn, implies that for the critical (tenth) maze, the larger the number of incorrect right-hand circuitous paths, the longer will *S* fail to see the correct direct path.

For decades the Einstellung effect has been treated in isolation from learning theory, primarily because that theory was based on conditioning principles. This deficiency vanishes when the effect is viewed from the perspective of *H* theory. The Einstellung-like results in the second set of experiments were predicted from the basic principles of *H* testing. By the straightforward development of *H* theory, one suddenly realized that the theory encompassed one form, at least, of the Einstellung effect: The *S* appears functionally blind as long as he is sampling from an incorrect *H* set. This easy, natural treatment of a hitherto isolated phenomenon further attests, along with the confirmed predictions, to the validity of the *H* model described above.

REFERENCES

BANDURA, A. *Principles of behavior modification.* New York: Holt, Rinehart & Winston, 1969.

BOURNE, J. E., JR., GUY, D. E., DODD, D. H., & JUSTESEN, D. R. Concept identification: The effects of varying length and informational components of the intertrial interval. *Journal of Experimental Psychology*, 1965, **69**, 624–629.

Bruner, J. S., Goodnow, J. J., & Austin, G. A. *A study of thinking.* New York: Wiley, 1965.

Dixon, P. W., & Oakes, S. F. Effects of intertrial activity on the relationship between awareness and verbal operant conditioning. *Journal of Experimental Psychology*, 1965, **69**, 152–157.

Dulaney, D. E., Jr. Hypotheses and habits in verbal "operant conditioning." *Journal of Abnormal and Social Psychology*, 1961, **63**, 251–263.

Dulaney, D. E., Jr. Awareness, rules, and propositional control: A confrontation with S–R behavior theory. In T. R. Dixon & D. L. Horton (Eds.), *Verbal behavior and general behavior theory.* Englewood Cliffs, N. J.: Prentice-Hall, 1968.

Erickson, J. R. Hypothesis sampling in concept identification. *Journal of Experimental Psychology*, 1968, **6**, 12–18.

Frankel, F., Levine, M., & Karpf, D. Human discrimination learning: A test of the blank-trials assumption. *Journal of Experimental Psychology*, 1970, **85**, 342–348.

Glassman, W. E. Unsolved- and insoluble-problem behavior. Unpublished master's thesis, State University of New York at Stony Brook, 1970.

Gregg, L. W., & Simon, H. A. Process models and stochastic theories of simple concept formation. *Journal of Mathematical Psychology*, 1967, **4**, 246–276.

Holstein, S. B., & Premack, D. On the different effects of random reinforcement and presolution reversal on human concept identification. *Journal of Experimental Psychology*, 1965, **70**, 335–337.

Kanfer, F. H. Verbal conditioning: A review of its current status. In T. R. Dixon & D. L. Horton (Eds.), *Verbal behavior and general behavior theory.* Englewood Cliffs, N. J.: Prentice-Hall, 1968.

Krasner, L. Verbal operant conditioning and awareness. In K. Salzinger & S. Salzinger (Eds.), *Research in verbal behavior and some neurophysiological implications.* New York: Academic Press, 1967.

Levine, M. A model of hypothesis behavior in discrimination learning set. *Psychological Review*, 1959, **66**, 353–366.

Levine, M. Cue neutralization: The effects of random reinforcements upon discrimination learning. *Journal of Experimental Psychology*, 1962, **63**, 438–443.

Levine, M. Mediating processes in humans at the outset of discrimination learning. *Psychological Review*, 1963, **70**, 254–276.

Levine, M. Hypothesis behavior by humans during discrimination learning. *Journal of Experimental Psychology*, 1966, **71**, 331–338.

Levine, M. Neo-noncontinuity theory. In G. H. Bower & J. T. Spence (Eds.), *The psychology of learning and motivation.* New York: Academic Press, 1969.

Levine, M., Miller, P., & Steinmeyer, C. H. The none-to-all theorem of human discrimination learning. *Journal of Experimental Psychology*, 1967, **73**, 568–573.

Levine, M., Yoder, R. M., Kleinberg, J., & Rosenberg, J. The presolution paradox in discrimination learning. *Journal of Experimental Psychology*, 1968, **77**, 602–608.

Luchins, A. S. Mechanization in problem solving: The effect of *Einstellung*. *Psychological Monographs*, 1942, **54**(6, Whole No. 248).

Mandler, G., Cowan, P. A., & Gold, C. Concept learning and probability matching. *Journal of Experimental Psychology*, 1964, **67**, 514–522.

Postman, L., & Sassenrath, J. The automatic action of verbal rewards and punishments. *Journal of General Psychology*, 1961, **65**, 109–136.

Ress, F. C. The effect of previous sequence problems on simple discrimination learning. Unpublished master's thesis, Indiana University, 1965.

Ress, F. C., & Levine, M. Einstellung during simple discrimination learning. *Psychonomic Science*, 1966, **4**, 77–78.

Restle, F. The selection of strategies in cue learning. *Psychological Review*, 1962, **69**, 329–343.

Restle, F. Grammatical analysis of the prediction of binary events. *Journal of Verbal Learning and Verbal Behavior*, 1967, **6**, 17–25.

Restle, F., & Emmerich, D. Memory in concept identification: Effects of giving several problems concurrently. *Journal of Experimental Psychology*, 1966, **71**, 794–799.

Silver, D. S., Saltz, E., & Modigliani, V. Awareness and hypothesis testing in concept and operant learning. *Journal of Experimental Psychology*, 1970, **84**, 198–203.

Spence, K. W. An experimental test of the continuity and noncontinuity theories of discrimination learning. *Journal of Experimental Psychology*, 1945, **35**, 253–266.

Spielberger, C. D., & DeNike, L. D. Descriptive behaviorism versus cognitive theory in verbal operant conditioning. *Psychological Review*, 1966, **73**, 306–326.

Trabasso, T., & Bower, G. H. Presolution dimensional shifts in concept identification: A test of the sampling with replacement axiom in all-or-none models. *Journal of Mathematical Psychology*, 1966, **3**, 163–173.

Trabasso, T., & Bower, G. H. *Attention in learning.* New York: Wiley, 1968.

Verplank, W. S. Unaware of where's awareness: Some verbal operants—notates, monents, and notants. In C. W. Eriksen (Ed.), *Behavior and awareness.* Durham, N. C.: Duke University Press, 1962.

(Received August 7, 1970)

Psychological Review
1971, Vol. 78, No. 2, 141–150

AN ANALYSIS OF TWO QUANTITATIVE THEORIES OF COGNITIVE BALANCE[1]

A. RODNEY WELLENS AND DONALD L. THISTLETHWAITE[2]

Vanderbilt University

Previous attempts to formulate balance theories capable of accommodating interelement relations of varying magnitudes are briefly reviewed. Two new algebraic formulations of the quantitative balance theories of Feather and Wiest are proposed. It is shown that the discrepancy principle used in the former theory contains an algebraic formulation which is equivalent to one that may be given for the upper boundaries of balance implied in the latter theory. Empirical results probing these theories provided some support for the algebraic formulation common to both theories, but disconfirmed the polarization principle which has been used as an adjunct to Wiest's geometric model.

An early attempt to incorporate the concept of "cognitive consistency" into workable psychological theory was Heider's (1946) theory of cognitive balance. Cartwright and Harary (1956), after proposing an extension of the Heider model, observed that a principle problem yet to be resolved was the development of a systematic method of treating interelement relationships of varying strength. Osgood and Tannenbaum's (1955) principle of congruity accommodated two relations that may vary in magnitude, but as Wiest (1965, p. 1) and Zajonc (1968, p. 354) have recently observed, the third relation (the assertion linking source and concept) assumed only a positive or a negative sign and was not sensitive to finer variation. Numerous writers have since called attention to the need for a quantitative formulation of balance theory. For example, Insko (1967, p. 174) noted that balance theory should be extended to allow for small graduations in both sentiment and unit relations; Shaw and Costanzo (1970, p. 193) observed that the magnitude of a relation would be expected to be an important determinant of balance. Morrissette (1958) proposed a formula for expressing the degree of balance in s-graphs containing lines (relations) varying in strength. However, this formulation is not applicable to s-digraphs containing a single semicycle, the paradigm most commonly examined in the literature. More recently, Wiest (1965) and Feather (1966) independently proposed models capable of treating all relationships in such triadic structures quantitatively while at the same time defining a state of balance. It appears, however, that these latter models also have limitations: Wiest's geometric model is cumbersome when implications other than the correlational relationships he discussed are to be deduced, and Feather's discrepancy principle leaves balance undefined when one relation has a neutral value. The present paper offers heretofore unavailable, algebraic formulations of the quantitative balance theories of Wiest and of Feather, discusses conditions under which the two formulations yield similar and dissimilar predictions, and briefly reports the results of an experiment which probes the adequacy of each formulation.

WIEST'S MODEL

Wiest (1965) has presented a quantitative extension of Heider's balance theory, applicable to three element configurations involving a person (P), another person (O), and an attitude object (X). The three interelement relations, P's attitude toward X (P → X), P's perception of O's attitude toward X (O → X), and P's liking for O (P → O), are treated as three independ-

---

[1] This work was supported in part by Grant GS-2658 from the National Science Foundation and by a grant from the National Institute of Mental Health (MH-08528).

[2] Requests for reprints should be sent to Donald L. Thistlethwaite, Department of Psychology, Vanderbilt University, Nashville, Tennessee 37203.

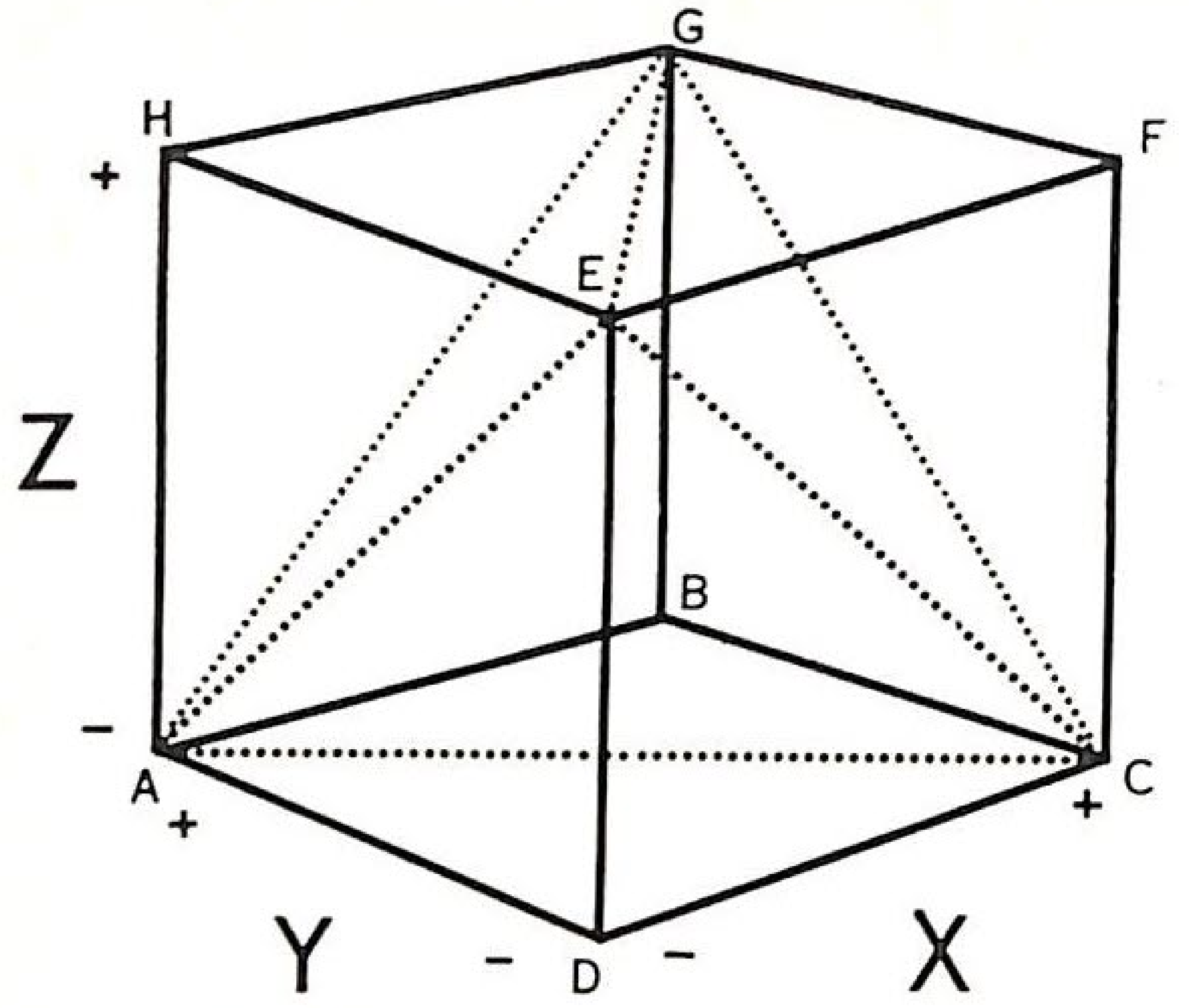


FIG. 1. Wiest's three-dimensional space as defined by the interelement relations P → X, O → X, and P → O, represented here by the axes of a cube *X, Y,* and *Z,* respectively.

ent bipolar continua which represent quantitative scales of the degree of attraction between pairs of elements. Figure 1 shows a cube whose three axes, *X, Y,* and *Z,* may be taken to represent these interelement continua. According to Wiest, a balanced triad is defined as any whose coordinates fall within or on the surface of the tetrahedron formed by connecting the four corners of the cube that correspond to the four balanced triads recognized by Heider (i.e., corners A, C, E, and G).

Wiest derived confirmable predictions from the tetrahedron model by "slicing" the cube along a single dimension at points which correspond to various magnitudes of a relation given in P-O-X configurations. These planes pass through the tetrahedron yielding "boundaries of balance" associated with triads possessing the given relational value. Such a procedure is illustrated in Section A of Figure 2 which shows the cube in Figure 1 sliced in five places (−2, −1, 0, +1, and +2) along the *Y* dimension. These boundaries yield predictions as to the magnitude of covariation expected between two of the relations as a function of the third. As can be seen, *X* and *Z* covary directly as a function of the level of *Y:* the stronger the *Y* relation, the greater the correlation between *X* and *Z.* The same pattern of boundaries shown in this figure would appear if the cube had been similarly cut along either of the other two dimensions.

Reformulation

It may be noted that the tetrahedron, ACEG, in Figure 1 has a top surface which includes the triangles CEG and AEG, while the bottom surface includes triangles ACE and ACG. When planes are passed through the cube at five equidistant points perpendicular to the *Y* axis (i.e., corresponding to the values −2, −1, 0, +1, and +2 along the *Y* axis in Figure 1) the intersection of each of these planes with the top surface of the tetrahedron results in the lines EC, ILK, MPO, RUT, and AG in Section A of Figure 2. These lines represent the upper boundaries of balance associated with triads possessing a given *Y* value of −2, −1, 0, +1, and +2, respectively. Similarly, the intersection of the five planes with the bottom surface of the tetrahedron is shown in Section A of Figure 2 as the lines EC, IJK, MNO, RST, and AG which represent the corresponding lower boundaries of balance associated with triads possessing the same given values.

The boundaries of balance discussed above define the theoretical upper and lower limits of *Z* for given values of *X* and *Y*. Because these boundaries are composed entirely of lines with slopes of plus or minus one, they may be expressed in terms of quite simple, linear equations. It can be shown (for example, by substitution of given *X* and *Y* values) that each of the predicted limits of *Z* in Section A of Figure 2 satisfies one of the following equations:

$$Z_U = k - |X - Y| \qquad [1]$$

$$Z_L = |X + Y| - k \qquad [2]$$

where Z_U represents the upper boundary limit for Z; Z_L represents the lower boundary limit for Z; $|X - Y|$ and $|X + Y|$ correspond to the absolute values of the real difference and sum, respectively, of the two given relations; and the integer k is a constant representing the highest positive scale value possible on each of the three interelement relations.

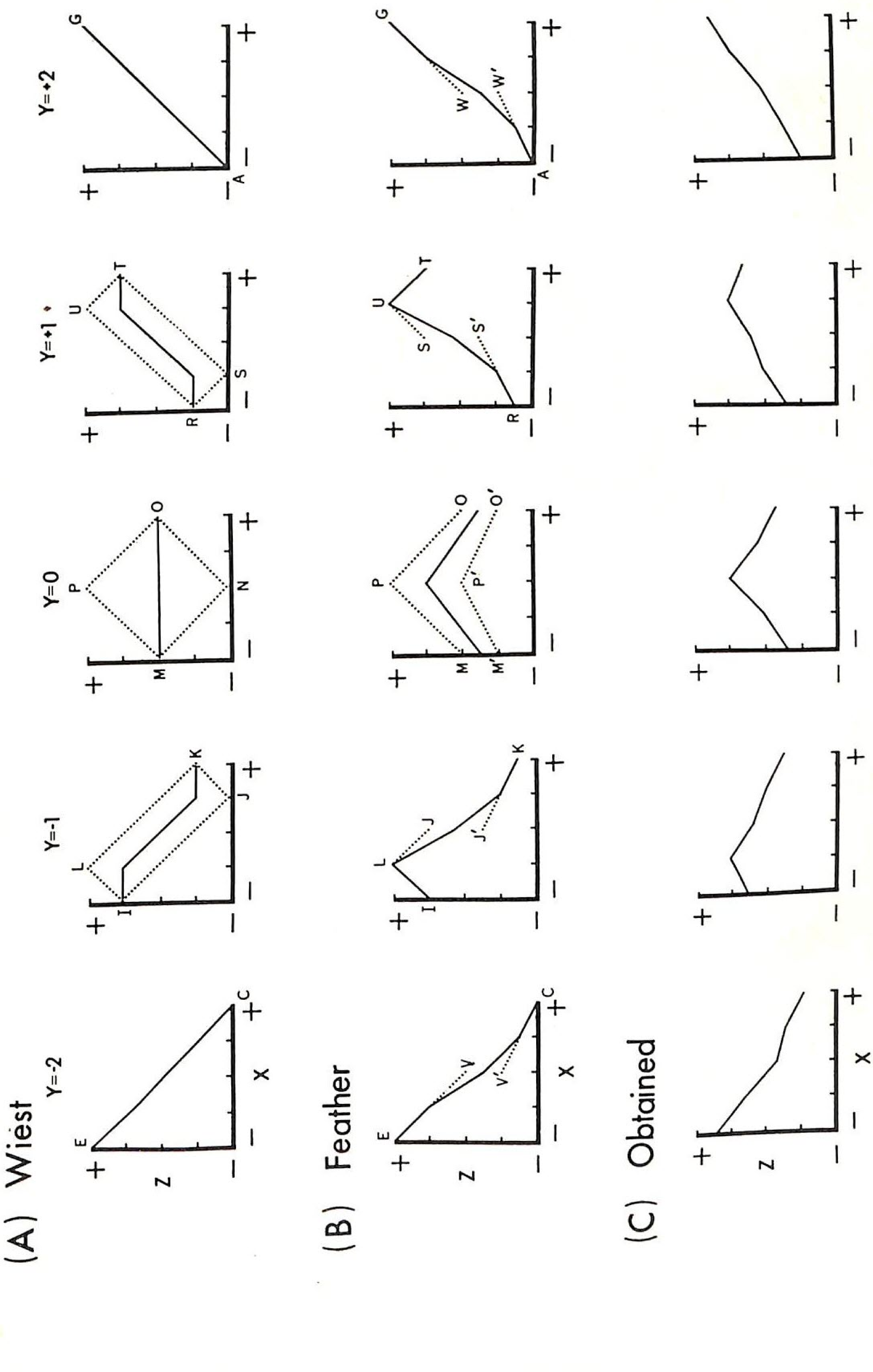


FIG. 2. Predicted and obtained mean ratings (*Z*) of subjects' estimated liking for another person given information concerning the other person's attitude (*Y*) toward an issue, and the subjects' attitude (*X*) toward the same issue.

Generalized Tetrahedron Formulation

The upper and lower limits of the range of values expected for Z, given values for relations X and Y, have been defined by the values Z_U and Z_L. These values may be combined to yield a generalized formula for predicting a single value for Z:

$$Z = aZ_U + bZ_L \qquad [3]$$

where Z represents the expected value of the missing relations, and a and b define weighting factors assigned to the range limits.

It should be noted that the values of a and b may vary according to assumptions made about the expected distribution of values along the range Z_L to Z_U. On the basis of Osgood and Tannenbaum's (1955) polarization principle, Wiest hypothesized a tendency for configurations to cluster toward the surfaces of the tetrahedron. If it is further assumed that these polarization effects tend to displace structures equally toward the surfaces of the tetrahedron, then it would seem appropriate to assign equal weights to the boundaries of balance associated with these surfaces. Predictions made according to this assumption may be construed as equivalent to assigning values of .5 to both a and b in Formula 3 and will hereafter be referred to as those based on the "equal weights" tetrahedron model. The solid lines in Section A of Figure 2 show mean predictions made according to this model for any missing relation in a triadic (P-O-X) structure given various values of two of the relations (the scale used in making these predictions ranged arbitrarily from -2 to $+2$).

One implication of the polarization assumption is that the obtained distributions of values of Z should be bimodal for all structures in which Z_L and Z_U do not coincide. It may be noted, however, that a variety of alternative distributions of Z values within the theoretical Z_L to Z_U ranges might be considered: for example, one might expect a normal distribution of values along any range, or a rectangular distribution of equally probable values, or perhaps even a skewed distribution with the bulk of the cases clustering about either the upper or lower boundaries of balance. In any case, any specific assumption regarding the predicted distribution of points inside the tetrahedron can be regarded as a dispensable adjunct to the generalized geometric model.

FEATHER'S MODEL

Feather's (1964) structural balance model of communication effects is based on Cartwright and Harary's (1956) reformulation of Heider's balance theory. Attitudinal relationships considered in Feather's model deal with a source (S) presenting a communication (C) about an issue (I) to a receiver (R). Only three of these elements will be of concern here (namely, S, I, and R), with relations being limited to the receiver's attitude toward an issue ($R \rightarrow I$), the receiver's perception of the source's attitude toward the issue ($S \rightarrow I$), and the receiver's attitude toward the source ($R \rightarrow S$). Recently, Feather (1966) has shown that the sign and relative magnitude of any one of these relations may be predicted when the sign and magnitude of the other two relationships are known or given. According to Feather, any given relation may be either positive or negative, and in addition, may be either weak or strong, resulting in a 4-point scale ranging from strong negative to strong positive. One of the principles of combination proposed by Feather may be called a discrepancy principle; it states that when two given attitudinal relations are of the same sign, small discrepancies in their magnitudes will be associated with (or will induce) a strong positive third relation, but, when two given attitudinal relations are of opposite signs, large discrepancies will be associated with a strong negative third relation.[3] A corollary which seems implicit in Feather's formulation is the following: Given two attitudinal relations of the same sign, the smaller the discrepancy in their magnitudes, the stronger will be the positive, third relation; given two attitudinal relations of opposite sign, the larger the discrepancy in their magnitudes, the stronger will be the negative, third relation.

---

[3] Although Feather (1966) stated that his proposed principles of combination were "in the nature of guesses based on one's intuitive knowledge of interpersonal relations," he also acknowledged that his discrepancy principle was suggested by Newcomb's (1959, 1961) formulation of the conditions for strain toward symmetry.

Extension and Reformulation

The scale which Feather proposes may be expanded to include any number of magnitudes. Such a scale may be conceived as a range of magnitudes extended from $-k$ to $+k$, where k may be any integer. Discrepancy may then be defined as the distance between any two values along this continuum. If X, Y, and Z are taken to represent the relations R → I, S → I, and R → S, then Feather's discrepancy principle may be rewritten as two functional relationships:

$$Z_S = k - |X - Y| \qquad [4]$$

$$Z_O = -\frac{|X - Y|}{2} \qquad [5]$$

Here Z_S or Z_O represents the missing (or to-be-induced) relational value; k represents the highest possible positive value on the scale employed; and $|X - Y|$ corresponds to the absolute value of the real difference between the two known relational values in the given structure. Formula 4 is used when the given values for X and Y are of the same sign. This formula yields a maximally positive value for Z_S when X and Y are minimally discrepant; Z_S approaches zero as $|X - Y|$ increases. The lines EV, ILJ, MPO, SUT, and WG in Section B of Figure 2 show predicted Z_S values based on this formula. Formula 5 is used when X and Y are of opposite signs and it yields the largest negative value for Z_O when X and Y are maximally discrepant; Z_O approaches zero as $|X - Y|$ decreases. The lines V′C, J′K, M′P′O′, RS′, and AW′ in Figure 2 show the predicted Z_O values based on this formula. It may be noted that when it is assumed that an attitude is perfectly juxtaposed between positively and negatively affective positions (i.e., when one holds, or is perceived to hold, an attitude that is located precisely at the "middle" position on the X or Y scale), both Formulas 4 and 5 may be considered applicable. Thus, for certain combinations of X and Y values, Formulas 4 and 5 yield competing predictions concerning the missing Z relation (cf. V vs. V′, J vs. J′, etc., in Section B of Figure 2). One alternative in using these formulas when X or Y has such middle values is to use the average of the competing Z_S and Z_O values as the prediction for the missing relation. The solid lines in Section B of Figure 2 show predictions made using this convention; these predictions will hereafter be referred to as those based on the "composite" discrepancy principle. Again, the scale used in making these predictions ranged from -2 to $+2$.

COMPARISONS BETWEEN THE TWO FORMULATIONS

The formulations presented here represent two different methods of combining information to arrive at independent predictions concerning missing attitudinal relations in analogous triadic situations. One point of similarity that exists between the Wiest and Feather models concerns a single formula common to both algebraic formulations. Formula 4 used in conjunction with combining values of the same sign in the extended Feather model is identical to Formula 1 used for finding the upper boundary limits in the reformulated Wiest model. It should be emphasized, however, that interpretations of the values obtained from these two equations vary significantly. Formula 4 in the Feather model yields a value for Z_S, which is taken as the predicted point for a missing relation computed on the basis of relations having the same sign; on the other hand, values for Z_U in the Wiest model define the upper limits for predictions made involving relations of any sign. Figure 2 shows that predictions for Z_S from Formula 4 (EV, ILJ, MPO, SUT, and WG) represent only a subset of the predictions for Z_U from Formula 1 (EC, ILK, MPO, RUT, and AG).

Examination of the solid lines presented in Sections A and B of Figure 2 shows the differential treatment of the $X \times Y$ interaction predicted by the equal-weights tetrahedron model and by the composite discrepancy model. When one of the relations given has a value which is low in magnitude (e.g., $Y = +1$, or -1), predictions made according to the former model approach an asymptote at plus or minus the lowest value given. Under the same conditions, predictions made according to the latter formulation tend to peak when given values are nondiscrepant, regardless of the strength of the relations involved.

Another interesting difference between the predictions of these two models concerns predictions made when one given relation has a magnitude of zero (e.g., $Y = 0$). Under these conditions, the equal-weights tetrahedron model predicts a magnitude of zero for the missing relation regardless of the strength of the remaining relation. The composite discrepancy model, on the other hand, predicts an inverted V relationship between the missing relation (Z) and the second relation (X). This is to say that when one relation in a P-O-X configuration is given as neutral, the degree of positivity or negativity of the missing relation will depend solely on the absolute strength of the third relation. As can be seen in Figure 2, when Y is given as neutral, Z increases in positivity as X decreases in absolute magnitude.

These two models become less discrete in their predictions when given relations take on extreme values. Thus, when one given value, say Y, is maximally positive, a strong positive correlation is expected between levels of the second given relation and the induced missing relation. Likewise, both models predict a strong negative correlation between levels of one given relation and the missing relation when the other given relation remains maximally negative. Although the equal-weights tetrahedron model and the composite discrepancy principle vary as to the degree of linearity associated with the expected relation between X and Z, it may be noted that the two theoretical curves are extremely similar when $Y = \pm 2$.

ILLUSTRATIVE APPLICATION OF THE ALGEBRAIC FORMULATIONS

There are, of course, numerous ways to investigate the implications of balance formulations. For example, Wiest (1965) attempted to measure "real" attitudes of children toward their classmates and found that the more a child liked another child, the more he tended to agree with the other in respect to the degree of his liking for a focal child. On the other hand, Aronson and Cope (1968) used a "completion" paradigm involving the direct induction of two "real" attitudes, with the dependent variable being the way the subject completed the third relation. Briefly, a given attitude toward an experimenter was first induced by having the experimenter act in a harsh or pleasant manner; next, an attempt was made to manipulate the subject's perception of the supervisor's attitude toward the experimenter by having the supervisor behave in a harsh or pleasant manner toward the experimenter within earshot of the subject; finally, the subject's acquired attitude toward the supervisor was assessed. Similarly, Fuller (1970) used the completion paradigm by employing the "bogus questionnaire" technique of Byrne (1969) to manipulate interpersonal similarity, thereby inducing affective relations toward two strangers (O and Q), and then asking subjects to predict O's probable liking for Q. More commonly, investigators have used "role-playing" instructions which ask the subject to imagine specified types of relationships. In a paradigm which may be called a "felt tension" variant, the subject is given three relations in a balanced or imbalanced triad and then rates the pleasantness of the hypothetical situation. In contrast, a completion paradigm using role-playing instructions typically gives the subject only two of the three interelement relations in a given triad and then asks him to estimate the third relation. Feather (1966), for example, has used the latter procedure in testing his formulation of the discrepancy principle.

The following experiment was designed to compare the ability of each of the given algebraic formulations to predict values associated with one particular interelement relation. Subjects were requested to imagine given P → X and O → X relationships and then asked to estimate their probable liking for an imagined other person (P → O).[4] Obviously, this experiment was not intended to widely sample or probe the relative applicability of these formulations, but serves mainly to illustrate one of several means of comparison.

Method

An attitude questionnaire was constructed with the stated purpose being "to investigate the processes by which people form first impressions of others on the basis of limited information." Each item of the 25-item questionnaire specified the subject's hypothetical stand on a "resolution about recent student demonstrations occurring on the West Coast" (P → X), together with another person's feeling toward the same resolution (O → X). Thus, for example, in varying the P → X relation, subject's were asked in successive items to imagine that they were "strongly opposed to," "mildly opposed to," "just as much in favor of as opposed to," "mildly in favor of," or "strongly in favor of" the hypothetical issue. These descriptive phrases may be thought of as representing the quantities -2, -1, 0, $+1$, and $+2$, respectively, on the X dimension of both the Wiest and Feather models. The O → X relation was similarly varied and may be conceived of as corresponding to identical values along the Y dimension. Subjects were instructed to mark, on a 7-point scale ranging from "dislike" to "like," how they thought they would feel toward the other person (P → O). Judgments were to be based solely on the information provided within a specific item.

The 25 items in this questionnaire represented the 25 cells of a Latin square formed by considering five levels each of the two independent variables, X and Y. The order of the items was determined randomly, with the restriction that no cells tangent to each other in the Latin square could be adjacently positioned as items on a given page of the questionnaire. This procedure was designed to prevent any sequential incrementing of the independent variables that would lead to *S*s responding according to an externally suggested pattern. Order was further controlled by creating a between-subject variable which varied the order in which pages were assembled; three such orders were prepared.

Subjects and procedure. A total of 150 students enrolled in an introductory psychology course at Vanderbilt University were asked to complete the questionnaire described above. Subjects were told to read the instructions printed on their questionnaire and to begin marking items as soon as they understood what they were to do. Any questions were answered individually. The session lasted approximately 40 minutes.

Results

This experiment represented a 5 × 5 × 3 factorial design, with repeated observations on the first two factors; these two within-subject factors represented the five levels of X and Y to which each subject was exposed in all possible combinations, while the third, between-subject, factor represented the three orders of item presentation. The dependent variable consisted of the subject's 7-point rating of his probable liking for the other person (Z). These ratings were transformed to a 5-point scale ranging from -2 to $+2$.

An analysis of variance of these ratings indicated that the overall interaction predicted by both the equal-weights tetrahedron model and the composite discrepancy principle was highly significant ($p < .001$, $\omega^2 = .46$). In addition, main effects were found to be associated with both X and Y ($p < .001$, $\omega^2 = .02$; $p < .001$, $\omega^2 = .02$). Although the formulation based on an equal-weights version of Wiest's model predicted no main effects, such effects were predicted according to the composite discrepancy model. As can be seen, if the Y classification is ignored in Section B of Figure 2, the latter model predicts that Z will have its largest value when $X = \pm 1$ and smallest value when $X = \pm 2$. A similar statement holds when the X classification is ignored and the main effects of Y on Z are considered. Generally the obtained effects conformed to these expectations. The third, between-subject, factor, representing the order of presentation of items, showed no significant main or interaction effects with the other two factors.

Examination of Section C of Figure 2 shows that the asymptotic relationship predicted according to the equal-weights tetrahedron formulation was not confirmed. Instead, when one of the relations given (e.g., Y) had a value of low magnitude, Z values peaked at nondiscrepant points as predicted according to the composite discrepancy model. Individual comparisons (Newman-Keuls) performed between these "peak" means and the other means associated with the particular level of Y involved showed both peaks to be significant ($p < .01$). Composite discrepancy predictions concerning given neutral values were also confirmed: when Y was given as zero, ratings of Z for $X = 0$ were significantly greater than ratings of Z for other values of X by a Newman-Keuls test ($p < .01$).

---

[4] The relations O → X, P → X, and P → O will be considered equivalent to the relations S → I, R → I, and R → S.

The general correlational predictions involving extreme values of Y were confirmed for both the equal-weights tetrahedron model and the composite discrepancy principle. When it was given that $Y = +2$, the correlation between given X values and obtained Z ratings was $+.98$; and when $Y = -2$, the correlation between these variables was $-.96$.

Distribution analyses performed on scores contributing to each of the 25 obtained means were examined for evidence of bimodality. In eight out of the nine cells in which the tetrahedron model yields Z_L to Z_U ranges greater than zero, the distributions of obtained Z ratings were unimodal (and each of the remaining 16 cells yielded a similar unimodal distribution of ratings). This finding is in direct conflict with the assumption of an overall polarization tendency operating within the tetrahedron, pushing structures toward its surfaces.

While the analysis of variance and individual comparison criteria indicate somewhat greater support for the composite discrepancy principle than for the equal-weights tetrahedron model, other criteria provided mixed results. Table 1 summarizes the average squared deviations of the obtained Z ratings and the ratings predicted by each of the algebraic equations: Formulas 1 and 2 represent the upper and lower boundaries of balance, respectively, of Wiest's model; Formula 3′ is an equal-weights version of the generalized tetrahedron model; Formulas 4 and 5 represent formulations of Feather's discrepancy principle for given relations of similar and opposite signs, respectively; finally, Formula 6 represents the composite discrepancy principle (i.e., predictions were based on either Formula 4 or 5, as appropriate, and on an average of Z_S and Z_O values when both equations may be construed as making competing predictions). The last column of Table 1 shows correlations between obtained ratings and the ratings predicted by each formula across all cells in which a given formula was applicable. In general, the average deviations and correlations in the upper half of Table 1 indicate

TABLE 1

Average Squared Deviations and Correlations of Obtained Means and Means Predicted by Formulas Based on the Feather and the Wiest Models

Prediction equation	Averages for given value of Y					Overall average	Correlation of obtained versus predicted ratings[a]
	−2	−1	0	+1	+2		
Formulas based on the Wiest model							
1	.31	.35	.60	.30	.29	.37	.95
2	.31	1.53	2.34	1.57	.29	1.21	.63
3′ (with $a = b = .5$)	.31	.34	.27	.34	.29	.31	.90
Formulas based on the Feather model							
4	.24	.49	.60	.44	.09	.39	.90
5	.66	.97	.49	.96	1.09	.79	.94
6[b]	.43	.65	.02	.61	.52	.45	.95

[a] Correlations for Formula 1, 2, 3′, and 6 are each based on 25 pairs of means; those for Formulas 4 and 5 are each based on 17 pairs of means.

[b] Formula 6 represents a composite, using Formulas 4, 5, and an average of the values provided by these equations where appropriate.

that the upper boundaries of balance of the tetrahedron model (Formula 1) yielded predictions more accurate than the lower boundaries (Formula 2), but no more accurate than an equal-weights model (Formula 3′). Also, the lower half of Table 1 shows that the discrepancy principle for given relations of similar signs (Formula 4) gave more accurate predictions, according to the least-squares criterion, than did the discrepancy principle for relations of opposite signs (Formula 5) or the composite discrepancy principle (Formula 6). The correlations for the latter three formulas showed only negligible differences, however. Judged by the criterion of squared residuals, the equal-weights tetrahedron model yielded a somewhat closer fit to the data than the composite discrepancy principle.[5]

DISCUSSION

It is apparent that the choice of criteria largely determines any inferred suitability of one model over the other. Although the composite discrepancy formulation better reflects obtained relational profiles in the cases tested ($Y = \pm 1$ and $Y = 0$), the equal-weights tetrahedron model provides predictions that are closer in terms of absolute deviations from obtained means. Results of the experiment therefore do not permit us to conclude that one of the quantitative balance formulations is clearly superior to the other. It seems more appropriate to consider which part of each formulation is better supported by the data, and to note how varied data collection procedures might affect the outcome of such comparisons.

First, it is clear that there is little support for any assumption of a polarization tendency pushing points in the interior of the tetrahedron toward the surfaces of the tetrahedron equally. The present results are consistent with those reported by Fuller (1970) in yielding no supportive evidence of such polarization tendencies. Little would be lost by discarding this assumption.

As noted, the upper boundaries of balance derived from the tetrahedron model seem to yield a decidedly better basis for predicting subjects' judgments than did the lower boundaries associated with this geometric figure. While the concept of a lower limit of balance did not, by itself, appear to provide a basis for accurate predictions of ratings, nonetheless it does seem useful in that it provides the basis for "compromise" estimates as suggested by Formula 3. For example, if the coefficients a and b in this formula are assumed to have weights of .75 and .25, respectively, a somewhat better fit of the data is obtained: the overall average squared deviation is .22 and the overall correlation of predicted and obtained ratings is .95. Further, it seems likely that the optimal a and b parameter values in Formula 3 will vary according to the type of relationship the subject is asked to estimate. Wellens and Thistlethwaite (in press) report some evidence suggesting that the deviations of obtained ratings from predictions based on the lower surface of the tetrahedron are smaller when the subject is asked to estimate P → X or O → X, rather than P → O. The latter findings may be interpreted as indicating that subjects prefer not to imagine that they dislike another.

Similarly, we can conclude that the discrepancy principle of Feather provides more accurate predictions when the two given relations have similar signs than when they have opposite signs. The latter condition is, of course, one in which qualitative balance theory would predict that the subject will complete the missing relationship negatively (by imagining that he dislikes the other person). If it had been assumed at the outset that subjects tend to avoid attributing negative liking relations toward others, then there would have been a basis for explaining the relatively inaccurate predictions of the discrepancy principle when the given relations are of opposite signs.

---

[5] The large number of subjects included meant that the power of each test (Lewis, 1960, p. 374) performed to evaluate the significance of discrepancies between obtained and predicted ratings was unusually great. Thus, it was found that the obtained data deviated significantly ($p < .01$) from the theoretical curves based on each of the six equations for each of the Y values shown in Table 1. Generally, the magnitudes of the F ratios for these significance tests varied directly with the average squared deviations shown in this table. For example, the poorest fit was obtained for Equation 2 in predicting Z values when $Y = 0$: the average squared deviation was 2.34 (Table 1) and $F = 768.95$, $df = 5/596$. The best fit was obtained for Equation 6 in predicting Z values when $Y = 0$: the average squared deviation was .02 (Table 1) and $F = 5.40$, $df = 5/596$.

In short, if it is true that subjects generally avoid attributing negative liking relations to others, then the conditions sampled in this experiment may not have been optimal for assessing the applicability of Formulas 2 and 5. Obviously, numerous and varied applications of these algebraic formulations will have to be evaluated before the relative utilities of the several formulas can be confidently assessed.

REFERENCES

ARONSON, E., & COPE, V. My enemy's enemy is my friend. *Journal of Personality and Social Psychology,* 1968, **8**, 8–12.

BYRNE, D. Attitudes and attraction. In L. Berkowitz (Ed.), *Advances in experimental social psychology.* Vol. 4. New York: Academic Press, 1969.

CARTWRIGHT, D., & HARARY, F. Structural balance: A generalization of Heider's theory. *Psychological Review,* 1956, **63**, 277–293.

FEATHER, N. T. A structural balance model of communication effects. *Psychological Review,* 1964, **71**, 291–313.

FEATHER, N. T. The prediction of interpersonal attraction. *Human Relations,* 1966, **19**, 213–237.

FULLER, C. H. An analysis of a quantitative theory of cognitive balance. Unpublished doctoral dissertation, Vanderbilt University, 1970.

HEIDER, F. Attitudes and cognitive organization. *Journal of Psychology,* 1946, **21**, 107–112.

INSKO, C. A. *Theories of attitude change.* New York: Appleton-Century-Crofts, 1967.

LEWIS, D. *Quantitative methods in psychology.* New York: McGraw-Hill, 1960.

MORRISSETTE, J. An experimental study of the theory of structural balance. *Human Relations,* 1958, **11**, 239–254.

NEWCOMB, T. M. Individual systems of orientation. In S. Koch (Ed.), *Psychology: A study of a science.* Vol. 3. New York: McGraw-Hill, 1959.

NEWCOMB, T. M. *The acquaintance process.* New York: Holt, Rinehart & Winston, 1961.

OSGOOD, C. E., & TANNENBAUM, P. H. The principle of congruity in the prediction of attitude change. *Psychological Review,* 1955, **62**, 42–55.

SHAW, M. E., & COSTANZO, P. R. *Theories of social psychology.* New York: McGraw-Hill, 1970.

WELLENS, A. R., & THISTLETHWAITE, D. L. A comparison of three theories of cognitive balance. *Journal of Personality and Social Psychology,* in press.

WIEST, W. M. A quantitative extension of Heider's theory of cognitive balance applied to interpersonal perception and self-esteem. *Psychological Monographs,* 1965, **79**(14, Whole No. 607).

ZAJONC, R. B. Cognitive theories in social psychology. In G. Lindzey & E. Aronson (Eds.), *The handbook of social psychology.* Vol. 1 (2nd ed.) Reading, Mass.: Addison-Wesley, 1968.

(Received August 13, 1970)

Psychological Review
1971, Vol. 78, No. 2, 151–169

CONJOINT-MEASUREMENT ANALYSIS OF COMPOSITION RULES IN PSYCHOLOGY[1]

DAVID H. KRANTZ[2]
University of Michigan

AND

AMOS TVERSKY
Hebrew University, Jerusalem

Composition rules are theories that describe the relationships among several measurable variables. Conjoint measurement provides methods for analyzing such rules using ordinal information only. This analysis is applied to a class of four composition rules in three variables [$A + P + U$, $(A + P)U$, $AP + U$, APU], which have been widely employed in different areas of psychology. It leads to the formulation of observable ordinal properties that can be used to test and diagnose which of the rules, if any, is appropriate for a given set of data.

Much of scientific endeavor can be described as an attempt to account for changes in some dependent variable (e.g., response strength, subjective value, kinetic energy) as a consequence of variations in one or more independent variables. To attain this goal, the dependent as well as the independent variables have to be measured, and the function (or rule) that relates them has to be specified. If the variables that enter the proposed functional relation can be measured independently (as is typically the case in physics), the validity of the composition principle can be tested directly. For example, to verify the law that the kinetic energy, E, of a body of mass m, moving at velocity v is given by $E = \frac{1}{2} mv^2$, one has to measure E, m, and v independently, and then demonstrate, by computation, that the equality holds within the limits of accuracy of the measurements.

Many psychological theories have also been formulated as composition principles that describe a functional relation among several variables. Hull (1952), for instance, proposed that the reaction potential, or the response strength, R, of an organism is a *multiplicative* function of its drive D, habit strength H, and incentive K. Thus, Hull postulated the relation $R = DHK$. Spence (1956), on the other hand, argued that response strength equals the product of habit strength with the sum of drive and incentive, or in symbols, $R = H(D + K)$. We call this the *distributive* rule. Additional examples of psychological composition rules are discussed below.

The major difficulty in the study of composition rules in psychology is that, in general, there is no adequate independent procedure for the measurement of the relevant variables. These psychological variables (e.g., drive, incentive) are typically characterized in terms of some appropriate physical measures (e.g., hours of deprivation, magnitude of reward). This characterization may be very useful for various experimental purposes, but it provides neither a satisfactory definition nor an adequate measurement procedure, because the psychological attribute and the physically controllable variable may not be related in the proper manner. It appears rather arbitrary to assume anything more than a monotonic relation between response strength and running speed, for example. Furthermore, the psychological attribute of response strength may be characterized in terms of different physical measures such as running speed or proportion of correct responses, and there is no reason to suppose that these measures are related to each other by a linear transformation. Hence, the best one can hope for is to find a monotonic relation between the controllable physical scale and the relevant psychological variable. But if the physical scales provide at most an ordinal measure of the psychological attributes, how can any psychological composition rule be established?

---

[1] This project has been supported by National Institutes of Health Grants MH 11504 and GM 1231, and National Science Foundation Grants GB 4947, GB 6782, and GB 8181, all to the University of Michigan. We wish to acknowledge helpful comments by Maya Bar-Hillel, C. H. Coombs, R. M. Dawes, and R. D. Luce.

[2] Requests for reprints should be sent to David H. Krantz, Department of Psychology, 330 Packard Road, University of Michigan, Ann Arbor, Michigan 48104.

An answer to this question is given by the conjoint-measurement approach, which attempts to solve the measurement and the composition problems simultaneously, by constructing measurement scales for the relevant variables so that the proposed composition principle is satisfied. The problem investigated in this approach, therefore, is not whether a specified functional relation holds among several (independently measured) variables, but rather whether there *exist* scales of measurement (for both the dependent and the independent variables) that satisfy the proposed composition rule. The key feature of the conjoint-measurement approach is that only the ordinal aspects of the data are required to be compatible with the proposed composition principle. Thus, in conjoint measurement, or for that matter in any fundamental measurement, one starts with an ordering of the dependent variable and investigates what properties this order should satisfy so that it can be represented numerically according to the proposed composition principle.

For concreteness, consider a factorial experiment in which the effects of drive, habit strength, and incentive are manipulated by varying, respectively, the length of the period of food deprivation, the amount of previous training in the maze, and the concentration of food reward. Suppose that response strength is ordered by the time it takes the organism to traverse the maze, and let $(d,h,k) > (d',h',k')$ denote the observation that organisms traverse the maze faster (hence show greater response strength) in treatment combination (d,h,k) than in treatment combination (d',h',k'), of drive, habit strength, and incentive, respectively. From the viewpoint of conjoint measurement, however, the actual speeds at which the maze is traversed are inessential—only their order matters. The conjoint-measurement analysis of the Hullian model consists, then, of the study of the ordinal conditions that are necessary and/or sufficient for the existence of scales D, H, and K for the measurement of drive, habit strength, and incentive, respectively, satisfying

$$(d,h,k) \geq (d',h',k') \text{ if and only if } D(d)H(h)K(k) \geq D(d')H(h')K(k').$$

Thus, the composition rule proposed by Hull is satisfied in an ordinal sense, if there *exist* scales D, H, and K whose product preserves the observed order of response strength. Similarly, the distributive composition rule, proposed by Spence, is satisfied in an ordinal sense, if there exist scales that satisfy the relation

$$(d,h,k) \geq (d',h',k') \text{ if and only if } [D(d) + K(k)]H(h) \geq [D(d') + K(k')]H(h').$$

Given this interpretation of the problem, the question becomes what are the qualitative (i.e., ordinal) conditions that the data must satisfy in order to be compatible with the composition rules proposed by Hull or Spence. Indeed, one may wonder whether it is possible to distinguish between these rules on the basis of ordinal information alone. It turns out, however, that the two composition rules are qualitatively different, and that the multiplicative rule, for example, implies testable ordinal properties that are not implied by the distributive rule. To illustrate, suppose $(d,h,k) > (d',h',k)$ and $(d',h',k') > (d,h,k')$. Such data violate the Hullian model, but not the Spencian model. Under the Hullian model, the first inequality implies $D(d)H(h)K(k) > D(d')H(h')K(k)$. Assuming that all the values are positive, we can multiply this inequality by $K(k')/K(k)$ to obtain $D(d)H(h)K(k') > D(d')H(h')K(k')$ which, under the Hullian model, implies $(d,h,k') > (d',h',k')$ contrary to our hypothesis. On the other hand, if we let $D(d) = 1$, $D(d') = K(k') = 2$, $K(k) = H(h') = 4$, and $H(h) = 5$, we see that the Spencian model is compatible with the data, because $(1 + 4)5 > (2 + 4)4$ and $(2 + 2)4 > (1 + 2)5$.

The multiplicative and the distributive models, therefore, can be compared on the basis of testable ordinal properties, when all three factors are varied simultaneously. If one or more factors are held constant, however, and the scales are all positive, then the two models can be shown to be ordinally equivalent, that is, they are indistinguishable on the basis of ordinal information. Unfortunately, in most of the experiments designed to compare the models, for example, Seward, Shea, and Davenport (1960) and Pavlik and Reynolds (1963), only two of the factors were varied. Hence, any conclusion based on such tests is necessarily limited to the particular measurement scales employed.

Viewed numerically, all composition rules are equally refutable. This is no longer true, however, when they are evaluated from an ordinal viewpoint. For example, assuming that all scale values are positive, it can be shown that more orderings are compatible with the distributive rule than with the multiplicative rule. The Hullian model, therefore, is easier to refute ordinally than the Spencian one. Such information must be taken into account in order to evaluate the models properly.

The search for ordinal properties that lead to the construction of numerical scales and to the formulation of composition rules has attracted much research both in the physical and the behavioral sciences. Around the beginning of the century, Hölder (1901) presented a set of assumptions, or axioms, formulated in terms of an ordering of objects with respect to some attribute, and showed that these qualitative assumptions yield the familiar additive measurement scale for attributes like mass, length, or time. Perhaps the earliest and the best known example of measurement theory in the behavioral sciences is the theory of expected utility that was first formalized by von Neumann and Morgenstern (1951), and later generalized by Savage (1954) and by Luce and Krantz (1971). This theory consists of a set of qualitative assumptions about an individual's preference order between uncertain options, or lotteries, that lead to the construction of utility and subjective-probability functions that satisfy the expected-utility principle.

Recent research on fundamental measurement originated with investigations of the additive rule (e.g., Adams & Fagot, 1959; Davidson, Suppes, & Siegel, 1957; Debreu, 1960; Luce & Tukey, 1964), and later extended to more complicated rules (e.g., Krantz, 1968; Roskies, 1965; Tversky, 1967a). This work, motivated primarily by theoretical problems in social science, has led to the development of new methods for the ordinal analysis of a wide variety of composition rules, and for the construction of numerical scales. For a detailed account of these developments see Krantz, Luce, Suppes, and Tversky (1971). In the present paper, we analyze from an ordinal viewpoint a particular class of simple composition rules in three variables. Our purpose is to exhibit some testable ordinal properties that can be used to diagnose which, if any, of the various composition rules is appropriate. Consequently, this paper is primarily concerned with diagnosis and testing rather than with axiomatization and scaling. It should be noted, however, that if the data are sufficiently regular, or finely graded (i.e., equally spaced, or dense), then the proposed diagnostic properties are not only necessary but also sufficient for the validity of the rules in question and for the construction of the appropriate numerical scales. These results were first obtained by Krantz (1968), and later developed in full by Krantz et al. (1971, Ch. 6 and 7).

This paper is devoted to the analysis of a class of composition rules in three variables, called simple polynomials. In general, this class is characterized as follows: (*a*) any single variable is a simple polynomial; (*b*) if two simple polynomials have no variables in common, then their sum and their product are also simple polynomials. More specifically, there are exactly four[3] simple polynomials in three variables; namely,

$A + P + U$ additive rule
$(A + P)U$ distributive rule[4]

---

[3] We ignore the distinctions obtained by permuting variables, since the diagnosis and testing problem for $(A + P)U$ is the same as that for $(A + U)P$.

[4] This rule is called distributive because the multiplicative factor U is distributive over the additive factors A and P. The reason for the selection of these letters will become clear later.

$AP + U$	dual-distributive rule[5]
APU	multiplicative rule.

An example of a composition rule in three variables that is not a simple polynomial is $AP + PU + AU$. The simple polynomials in three variables are of special interest. First, they encompass a large class of psychological theories. Second, they are particularly tractable from a mathematical viewpoint, although their ordinal analysis is by no means trivial. Third, once the ordinal analysis of this class is performed, it can be extended to simple polynomials in any number of variables (see Krantz, 1968). To motivate the present investigation, we first list several examples of theories, from various areas of psychology, that are based on the above composition rules. Most of the theories, however, have been traditionally interpreted within a numerical, rather than an ordinal, framework. The present results, therefore, provide more general (i.e., ordinal) methods for testing and comparing such theories.

Examples

1. We have already seen that the performance models of Hull (1952) and Spence (1956) consist, respectively, of the multiplicative and the distributive composition rules. These models can be further extended by subtracting off a term attributable to an inhibition factor. This generalization leads to the polynomials $HDK - I$ and $(D + K)H - I$. If, say, K is held constant, then both models become ordinally equivalent to the dual-distributive rule $HD - I$. Clearly, subtraction is equivalent to the addition of a variable with a negative sign.

2. The additive composition rule has been applied by scores of authors, in many different areas of psychology, in either an ordinal or a numerical form. We cite just a few examples: Adams and Fagot (1959) and Tversky (1967b) in decision making; Hammond (1955) and Hoffman (1960) in clinical evaluation; Anderson (1970), Anderson and Fishbein (1965), and Tversky and Krantz (1969) in impression formation.

3. In his theory of achievement motivation, Atkinson (1957) postulated that the attractiveness of a risky task is given by $I_SP_SM_S - I_FP_FM_F$ where S and F denote success and failure, respectively, I the incentives of these states, P the probabilities of their occurrence, and M the strengths of the motives to achieve success and to avoid failure. In its general form, this is a simple polynomial in six variables. For situations where incentive is inversely related to probability, Atkinson has simplified the model by assuming that $I_S = 1 - P_S = P_F$, and $I_F = 1 - P_F$. Letting $Q = I_SP_S = I_FP_F$ yields the distributive rule $(M_S - M_F)Q$.

4. Several ordinal composition rules in three variables are obtained by applying expected-utility theory to specific situations. In particular, consider the lottery where one receives a certain amount of Commodity 1 and a certain amount of Commodity 2 if some chance event occurs, and nothing otherwise. Assuming that the two commodities are independent, the expected utility of such a lottery is given by the distributive rule $(U_1 + U_2)P$, where P is a (subjective) probability function and U_1, U_2 are the utilities of the two commodities.

5. Consider a game in which a fair coin is tossed several times, and where a person wins a certain prize each time the coin turns up heads, and pays a certain fine every time the coin turns up tails. Coombs and Huang (1970) have proposed that the risk of such a game is given by the distributive rule $(R + V)N$, where R is a function of the range of the outcomes, V is a function of the expected value, and N is a function of the number of repeated plays.

6. Another example of a distributive composition rule is derivable from the opponent-colors theory proposed by Hurvich and Jameson (1955). According to their theory, the saturation, S, of a color is expressible by

$$\frac{S}{1 - S} = (|r - g| + |y - b|)\frac{1}{|w - bk|},$$

where $|w - bk|$ is the absolute value of the achromatic (white–black) response, and $|r - g|$, $|y - b|$ are the absolute values of the chromatic red–green and yellow–blue responses, respectively.

7. In the models discussed thus far, negative scale values need not appear in a multiplicative form. There are some instances, however, in which one factor has a natural zero point, so that the values on the opposite sides of this point produce opposite effects on the ordering of another factor. In this case the use of negative scale values is suggested. For example, suppose that combinations of adjectives (e.g., helpful, tall, offensive) and adverbs of quantity (e.g., very, slightly) are evaluated with respect to moral connotation (e.g., good to bad). In such tasks, we find (for examples see Cliff, 1959; Gollob, 1968) that "very helpful" is better than "slightly helpful," "very offensive" is worse than "slightly offensive," and that "very tall" and "slightly tall" are both neutral, and hence equal, with respect to moral connotation. In this case, a multiplicative composition rule is called for, where the values of the adverbs are all positive, and the values of the adjectives helpful, tall, and offensive are, respectively, positive, zero, and negative. Another example of a composition rule based on both positive and negative values can be found in Rosenberg (1956).

8. Another composition rule that accommodates both positive and negative values is obtained by quantifying the theory of structural balance, pro-

---

[5] This rule is called dual distributive because it is obtained from the distributive rule by interchanging the roles of addition and multiplication.

posed by Heider (1946), which has been widely applied to the study of cognitive organization and interpersonal relations (for examples see Brown, 1965, Ch. 11).

Consider a group of three individuals, with various degrees of liking among its members, and assume, for simplicity, that liking (and disliking) is a symmetric relation. Following Heider's ideas, we propose that the degree of balance of the group is expressible by the multiplicative rule $L_{12}L_{23}L_{13}$, where L_{ij} is the positive (or negative) degree of liking (or disliking) between Individuals i and j. Thus, a group is balanced if either all its members like each other—"the friend of my friend is my friend"—or if one of its members dislikes the other two, who in turn like each other—"the friend of my enemy is my enemy." Likewise, a group is unbalanced if either all its members dislike each other—"the enemy of my enemy is my enemy"—or if one of its members likes the other two, who in turn dislike each other—"the enemy of my friend is my friend."

DIAGNOSIS AND TESTING

In this section we formulate several testable ordinal conditions and show how they can be used to diagnose and test simple polynomials in three variables. To simplify the exposition, we first discuss the *positive case* where all scale values are positive, and then investigate the more general *signed case* where zero and negative values are required. We use the boldface letters **A**, **P**, and **U** to denote variables, or factors, and the capital italicized letters A, P, and U to denote the scales that measure their respective effects. Note that **A**, **P**, and **U** are sets of treatment conditions, whereas A, P, and U are numerical functions defined on these sets. For notational convenience, we use letters from the beginning (e.g., a,b,c), the middle (e.g., p,q,r), and the end (e.g., u, v,w) of the alphabet to denote, respectively, elements of **A**, **P**, and **U**. Thus, $A(a)$ is the scale value of the element a in **A**, etc.

Independence

One class of simple testable properties, called independence conditions, plays a central role in the ordinal analysis of composition rules. The essence of ordinal independence is that the ordering of the dependent variable can be used to order some of the independent variables (i.e., factors) in a manner that does not depend on the remaining variables.

Suppose, for example, that $(a,p,u) \geq (b,p,u)$. Since the two treatment conditions differ only with respect to their first component, one is tempted to conclude that the effect of a is at least as large as the effect of b. This conclusion, however, is valid only if the former effect *always* exceeds the latter when the other components are fixed; that is, if for any q and v, $(a,q,v) \geq (b,q,v)$. Otherwise, the order of the elements of **A** is not well defined because it depends on the elements of **P** and **U** with which they are combined. Formally, **A** is *independent* (of **P** and **U**) whenever

$$(a,p,u) \geq (b,p,u) \text{ if and only if } (a,q,v) \geq (b,q,v). \quad [1]$$

(In stating formal properties of this type, it is understood that the quantification is universal. That is, **A** is independent whenever Equation 1 holds for *all* a, b in **A**; p, q in **P**; and u, v in **U**.) Independence for **P** (or for **U**) is defined in the same fashion.

The concept of ordinal independence can be applied not only to a single factor but to a pair of factors as well. Suppose, for example, that $(a,p,u) \geq (b,q,u)$. Since these two treatment combinations have the same third component, one may be tempted to conclude that the effect of the (a,p) combination is at least as large as that of the (b,q) combination. For this conclusion to be valid, however, the same order of the joint effects of **A** and **P** must hold for any fixed third component, that is, $(a,p,v) \geq (b,q,v)$. Stated formally, **A** and **P** are *jointly independent* (of **U**) whenever

$$(a,p,u) \geq (b,q,u) \text{ if and only if } (a,p,v) \geq (b,q,v). \quad [2]$$

Joint independence for **P** and **U** and for **A** and **U** is defined in similar fashion. (Note that joint independence for all pairs of factors implies independence for all single factors.[6] The converse does not hold in general.)

---

[6] To demonstrate, assume joint independence, and suppose that $(a,p,u) \geq (b,p,u)$. Since **A** and **P** are jointly independent, this inequality is equivalent to $(a,p,v) \geq (b,p,v)$ which, in turn, is equivalent to $(a,q,v) \geq (b,q,v)$ since **A** and **U** are jointly independent. Hence, $(a,p,u) \geq (b,p,u)$ is equivalent to $(a,q,v) \geq (b,q,v)$, and independence is thus derived from joint independence.

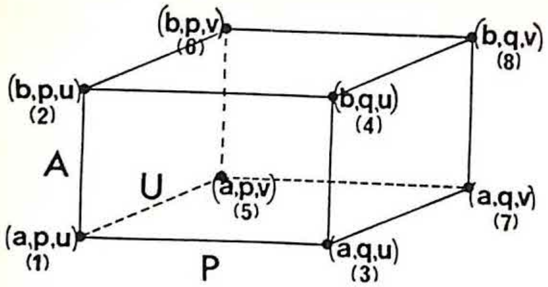


FIG. 1. A graphic illustration of a $2 \times 2 \times 2$ factorial structure.

A graphic illustration may elucidate the meaning of independence. Consider the $2 \times 2 \times 2$ factorial structure presented in Figure 1, whose vertices represent treatment combinations that are labeled by the appropriate letters and numbered from 1 to 8.

Using the rectangular solid displayed in Figure 1, independence and joint independence can be interpreted geometrically. Independence for **A**, for instance, corresponds to the property that endpoints of edges parallel to the **A** axis are ordered in the same way. For example $(2) \geq (1)$ if and only if $(8) \geq (7)$. More generally, single-factor independence corresponds to the condition that endpoints of parallel edges are ordered in the same fashion, for example, $(4) \geq (2)$ if and only if $(7) \geq (5)$.

Joint independence for **A** and **P** corresponds to a somewhat different property, namely, that pairs of vertices with the same **U** component, which lie on parallel lines, are ordered in the same way, for example, $(4) \geq (1)$ if and only if $(8) \geq (5)$. More generally, joint independence corresponds to the condition that endpoints of parallel diagonals are ordered in the same way.

Note that both single-factor and joint independence conditions are formulated in terms of a $2 \times 2 \times 2$ subset of a factorial structure. In order to test them, therefore, one has to consider all $2 \times 2 \times 2$ subsets of any larger three-factor experiment, and to check whether the required properties are satisfied within each such subset.

In practice, such tests of independence are quite simple. Given data from an $l \times m \times n$ three-factor experiment, one looks at the ordering of any "row" of l cells, in which the **A** component varies, with the **P** and **U** components fixed. All such orderings need to be identical (in error-free data), so that the **A** components can be permuted such that the order along any **A** "row" is monotone increasing. Similar tests of independence hold for **P** and for **U**. To test joint independence of **A** and **P**, one must check whether the ordering is the same in each **A** × **P** matrix, with the **U** component fixed, etc.

The first step in the diagnosis of composition rules is the test of single-factor independence. If it fails, for any factor, then no composition rule with positive scales is compatible with the data. In other words, if single-factor independence does not hold, then either no composition rule is available, or the composition rule must include some zero or negative scale values.

If single-factor independence is satisfied, then neither negative nor zero values are required. That is, if any composition rule is appropriate, positive scales suffice. Since this is the case in many applications, and since the introduction of zero and negative values results in certain complications, we first discuss the positive case (i.e., where simple independence holds), and treat the signed case in a later section. (Note that in the positive case, the additive and the multiplicative rules are (ordinally) equivalent, because the former can be obtained from the latter by a logarithmic transformation. In the signed case, where the restriction to positive values is not imposed, the additive model is equivalent to that special case of the multiplicative model in which all scale values are positive.)

Cancellation

Besides independence, there is another class of ordinal properties—called cancellation conditions—that are used to diagnose and test the various composition rules. The first cancellation condition presented is a necessary condition for all composition rules in the positive case. It is a two-dimensional property, that is, it applies to pairs of treatment combinations that have a common component in one factor. Formally, **A** and **P** satisfy *double cancellation*[7] if

$$(a,q,u) \geq (b,r,u)$$

and

$$(b,p,u) \geq (c,q,u), \qquad [3]$$

imply

$$(a,p,u) \geq (c,r,u).$$

Double cancellation is the key property necessary for additive composition in two variables (e.g., see Krantz et al., 1971; Luce & Tukey, 1964). A graphical illustration of double cancellation is presented in Figure 2, where treatment combinations are represented by points, and the inequalities of the hypotheses and the conclusion are denoted, respectively, by thin and thick arrows.

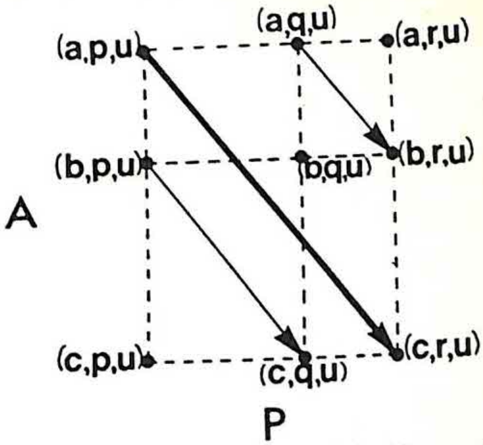


FIG. 2. A graphic illustration of double cancellation. (The inequalities of the hypotheses and the conclusion are represented, respectively, by thin and thick arrows.)

Unlike the independence conditions, which can be tested with only two levels per factor, double cancellation requires at least a 3 × 3 design. (As Figure 2 shows, six cells of a 3 × 3 matrix are involved.) In a larger, $m \times n$ design, there are $\binom{m}{3} \times \binom{n}{3}$ possible tests of double cancellation. Each of these can be classified in one of three ways: *acceptance* (the antecedents and the conclusion of Equation 3 hold), *rejection* (the antecedents hold, but the conclusion does not), or *no test* (the antecedents do not hold). The tests are not logically independent, since the conclusion of one test may be the antecedent of another, etc.

If independence holds but double cancellation is violated, then all simple polynomials are rejected. This follows from the fact that in all simple polynomials, any two-factor subset of treatment combinations, obtained by selecting a fixed level on the third factor, satisfies additivity, and hence satisfies double cancellation. In the positive case, therefore, both independence and double cancellation are necessary conditions for any simple polynomial.[8] Once these properites, that are common to all the rules, are satisfied, we can turn to the investigation of those properties that distinguish among the three composition rules.

We start by checking the number of pairs of factors that are jointly independent. If the additive (or, equivalently, the multiplicative) rule is satisfied, then all pairs of factors satisfy joint independence. In the other two rules [i.e., $(A + P)U$ and $AP + U$) **A** and **P** are jointly independent, but the other two pairs of factors are not. Hence, if no pair of factors is jointly independent, then no simple polynomial is appropriate; if one pair of factors is jointly independent, then both the distributive and the dual-distributive rules may be appropriate; and if all three pairs of factors are jointly independent, then the additive rule is appropriate.[9]

---

[7] The term cancellation is due to the fact that the conclusion can be obtained from the hypotheses by "cancelling" the common elements b and q on both sides of the inequalities.

[8] See also Adams and Fagot (1959).

[9] The numerical example in **D**, **H**, **K** near the beginning of the paper shows that **D** and **H** need not be jointly independent for the distributive rule $(D + K)H$ to hold. A similar example can easily be constructed for the dual-distributive rule. On the other hand, these violations of joint independence depend on the fact that the ranges of the scales are sufficiently large. Thus, in some circumstances, the distributive or dual-distributive rule might hold, yet due to narrow range of scale values, two or three pairs of factors might satisfy joint independence. If two pairs satisfy joint independence, then the additive rule is eliminated and one can proceed as though only one pair satisfied joint independence, anticipating, however, a constricted range. If all three pairs satisfy joint independence, the distributive and dual-distributive rules are not necessarily eliminated, but the additive rule is by far the most likely to work. Moreover, if the data are sufficiently finely graded, and all three pairs satisfy joint independence, then it can be shown that the additive rule is bound to work (Krantz et al., 1971).

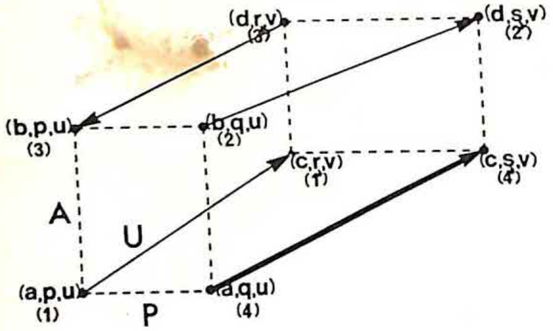


FIG. 3. A graphic illustration of distributive cancellation. (The inequalities of the hypotheses and the conclusion are represented, respectively, by thin and thick arrows.)

To complete the analysis of the positive case, the distributive and the dual-distributive rules have to be differentiated. For this purpose, we introduce two additional properties—distributive cancellation and dual-distributive cancellation. Each of these properties is satisfied by the respective composition rule and need not be satisfied by the other. Hence, they can be used to test which of the two rules, if either, is more likely to hold. (Both cancellation conditions, however, are satisfied by the additive composition rule.)

Distributive cancellation is said to be satisfied whenever

$$(a,p,u) \geq (c,r,v),\ (b,q,u) \geq (d,s,v),$$

and

$$(d,r,v) \geq (b,p,u) \qquad [4]$$

imply

$$(a,q,u) \geq (c,s,v).$$

A graphic illustration of this property is presented in Figure 3, where treatment combinations are represented by points that are both labeled and numbered. The inequalities of the hypotheses and the conclusion are denoted, respectively, by thin and thick arrows.

Interpreted in terms of Figure 3, distributive cancellation asserts that if two endpoints of a diagonal in the unprimed plane [e.g., (1) and (2)] are respectively larger than the two endpoints of a diagonal in the primed plane [e.g., (1′) and (2′)], then the two endpoints of the second diagonal in the primed plane [e.g., (3′) and (4′)] cannot be both respectively larger than the two endpoints of the second diagonal in the unprimed plane. In short, $(1) \geq (1')$, $(2) \geq (2')$, and $(3') \geq (3)$ imply $(4) \geq (4')$.

It is instructive to see how this property follows from the distributive rule. Let $A(a)$, $P(p)$, and $U(u)$ be the scale values of a, p, and u, respectively. Suppose the distributive rule holds; hence, from the hypotheses of distributive cancellation

$(a,p,u) \geq (c,r,v)$ implies

$$[A(a) + P(p)]\,U(u) \geq [A(c) + P(r)]\,U(v),$$

$(b,q,u) \geq (d,s,v)$ implies

$$[A(b) + P(q)]\,U(u) \geq [A(d) + P(s)]\,U(v),$$

and

$(d,r,v) \geq (b,p,u)$ implies

$$[A(d) + P(r)]\,U(v) \geq [A(b) + P(p)]\,U(u).$$

Adding the first two numerical inequalities yields

$$[A(a) + P(p) + A(b) + P(q)]\,U(u) \geq [A(c) + P(r) + A(d) + P(s)]\,U(v).$$

Subtracting the third numerical inequality (that follows from the third hypothesis) from this latest inequality yields

$$[A(a) + P(q)]\,U(u) \geq [A(c) + P(s)]\,U(v),$$

which, under the distributive rule, implies $(a,q,u) \geq (c,s,v)$. This establishes the necessity of distributive cancellation.

Finally, to test the dual-distributive rule, a more complicated ordinal property, that includes five hypotheses, is required. *Dual-distributive cancellation* is satisfied whenever

$$(c,s,v) \geq (a,r,w),\ (d,p,u) \geq (b,t,x),$$
$$(d,r,x) \geq (e,s,u),\ (e,q,y) \geq (c,t,y),$$

and

$$(a,t,y) \geq (d,q,y) \qquad [5]$$

imply

$$(a,p,v) \geq (b,q,w).$$

The dual-distributive cancellation is similar, in nature, to the other cancellation conditions, except that it is more difficult to test and interpret. The interested reader is invited to verify that this property follows from the dual-distributive composition rule.

A flowchart summarizing the procedure for the diagnosis of three-factor simple polynomials in the positive case is presented in Figure 4.

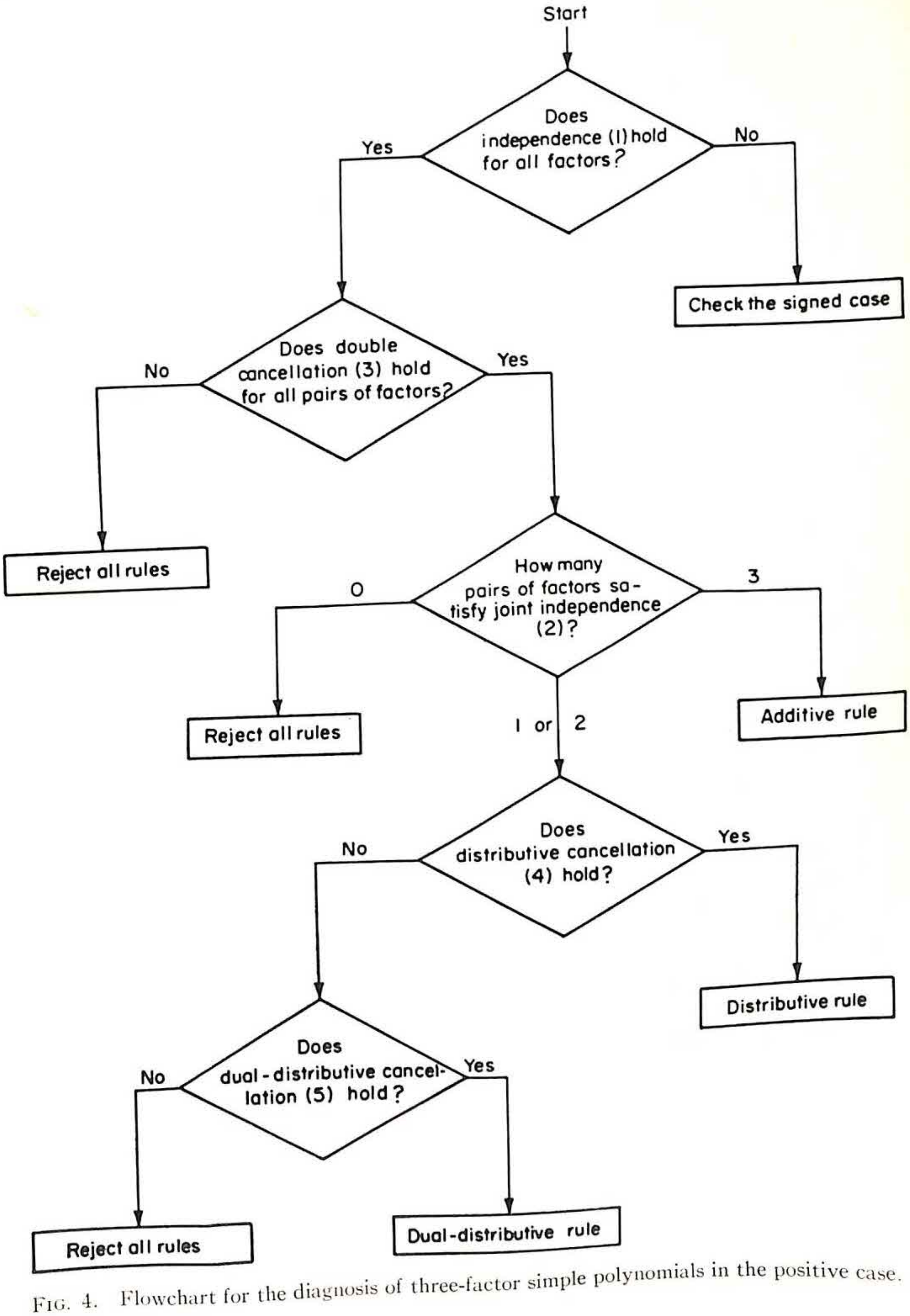

Start
Does independence (I) hold for all factors?
Yes
No
Check the signed case
Does double cancellation (3) hold for all pairs of factors?
No
Yes
Reject all rules
How many pairs of factors satisfy joint independence (2)?
0
3
Reject all rules
1 or 2
Additive rule
Does distributive cancellation (4) hold?
No
Yes
Distributive rule
Does dual-distributive cancellation (5) hold?
No
Yes
Reject all rules
Dual-distributive rule

FIG. 4. Flowchart for the diagnosis of three-factor simple polynomials in the positive case.

Sign Dependence

The diagnostic properties discussed in the previous section are all valid when the respective composition rules are restricted to positive-valued scales. If, however, the multiplicative factors admit negative or zero scale values, then these properties are readily violated. To demonstrate how independence (Equation 1) fails to hold, suppose treatment combination (p,u) has a "positive" effect, while treatment combination (q,v) has a "negative" effect, and assume $(a,p,u) > (b,p,u)$. According to the multiplicative composition rule, therefore, $(b,q,v) > (a,q,v)$, because a change in the sign of a multiplicative factor reverses the direction of the inequality. This violation of independence is due to the fact that treatment combinations (p,u) and (q,v) produce opposite effects or have different "signs." However, if we restrict the test to treatment combinations that have the same "sign," then independence is satisfied under all three composition rules. To accommodate negative and zero scale values, therefore, independence is generalized by taking "signs" into account. This can be accomplished if the set of all treatment combinations of the form (p,u) can be divided into three sets whose effects are, respectively, "positive," "neutral," and "negative." Specifically, **A** satisfies *sign dependence* if $\mathbf{P} \times \mathbf{U}$ can be partitioned into three sets, denoted $|\mathbf{P} \times \mathbf{U}|^+$, $|\mathbf{P} \times \mathbf{U}|^0$, and $|\mathbf{P} \times \mathbf{U}|^-$ such that **A** is independent of each of these sets (in the sense of Equation 1) and such that the positive and the negative sets induce opposite orders, and the neutral set induces an equality over **A**. More formally, let the pairs $|p,u|^+$, $|p,u|^0$, and $|p,u|^-$ denote, respectively, generic elements of $|\mathbf{P} \times \mathbf{U}|^+$, $|\mathbf{P} \times \mathbf{U}|^0$, and $|\mathbf{P} \times \mathbf{U}|^-$. The factor **A** is called *sign dependent* (on $\mathbf{P} \times \mathbf{U}$) whenever[10]

1. $(a,|p,u|^+) \geq (b,|p,u|^+)$ if and only if $(a,|q,v|^+) \geq (b,|q,v|^+)$.
2. $(a,|p,u|^-) \geq (b,|p,u|^-)$ if and only if $(a,|q,v|^-) \geq (b,|q,v|^-)$. [6]
3. $(a,|p,u|^+) \geq (b,|p,u|^+)$ if and only if $(b,|q,v|^-) \geq (a,|q,v|^-)$.
4. $(a,|p,u|^0) = (b,|p,u|^0)$.

Thus, **A** is sign dependent if the ordering induced over **A** by (p,u) depends only on the "sign" of (p,u), that is, only on the subset to which (p,u) belongs, with different subsets behaving according to their respective "signs." Clearly, independence (Equation 1) is the special case of sign dependence (Equation 6) where the sets of neutral and negative treatment combinations (i.e., $|\mathbf{P} \times \mathbf{U}|^0$, and $|\mathbf{P} \times \mathbf{U}|^-$) are both empty. In this case, Part 1 above reduces to Equation 1 while the other parts of Equation 6 are satisfied vacuously.

In terms of an $l \times m \times n$ factorial design, sign dependence for **A** means that the **A** components can be permuted so that the entries in the l cells of any one **A** "row" are either monotone increasing or monotone decreasing or constant. The fixed $\mathbf{P} \times \mathbf{U}$ elements have opposite "signs" when the corresponding **A** "rows" are oppositely ordered, or neutral "sign" when the **A** "row" is constant. If any two nonconstant **A** "rows" are ordered differently, but not oppositely, both independence and sign dependence fail for **A**.

Joint sign dependence for a pair of factors is defined in an analogous manner, except that only a single factor, the third one, is partitioned into signed subsets. Formally, **A** and **P** are said to be *jointly sign dependent* (on **U**) if **U** can be partitioned into three sets, denoted $\mathbf{U}^+$, $\mathbf{U}^0$, and $\mathbf{U}^-$, with respective generic elements u^+, u^0, and u^-, such that

1. $(a,p,u^+) \geq (b,q,u^+)$ if and only if $(a,p,v^+) \geq (b,q,v^+)$.
2. $(a,p,u^-) \geq (b,q,u^-)$ if and only if $(a,p,v^-) \geq (b,q,v^-)$. [7]
3. $(a,p,u^+) \geq (b,q,u^+)$ if and only if $(b,q,u^-) \geq (a,p,u^-)$.
4. $(a,p,u^0) = (b,q,v^0)$.

Thus, **A** and **P** are jointly sign dependent on **U** if they are jointly independent of $\mathbf{U}^+$, $\mathbf{U}^0$, and $\mathbf{U}^-$, and if the order induced by the positive set is the opposite of that induced by the negative set, and the order induced by the neutral set is an equality. Clearly, joint independence is the special case of joint sign dependence where the sets of neutral and negative treatment conditions (i.e., $\mathbf{U}^0$ and $\mathbf{U}^-$) are empty. In this case, Part 1 of Equation 7 reduces to Equation 2, while the other parts of Equation 7 are satisfied vacuously.

---

[10] Formally, 3 implies 1 and 2, but this logical nicety is not incorporated in Equation 6, since it may be confusing at first glance.

Joint sign dependence for **A** and **P** means that any two nonconstant $\mathbf{A} \times \mathbf{P}$ matrices (for different fixed components in **U**) are ordered either alike or entirely oppositely. If the latter, then the corresponding components in U have opposite sign. A constant $\mathbf{A} \times \mathbf{P}$ matrix is diagnostic of u in $\mathbf{U}^0$.

Although it is obvious how to designate $\mathbf{U}^0$, the $+$ and $-$ signs can be specified arbitrarily for any one partition. Other partitions, however, must have signs designated in a consistent manner. For example, in the case of the multiplicative composition rule, sign dependence for single factors yields partitions of $\mathbf{A} \times \mathbf{P}$, $\mathbf{A} \times \mathbf{U}$, and $\mathbf{P} \times \mathbf{U}$, while joint sign dependence yields partitions of **A**, **P**, and **U**. We must have the properties

$$|\mathbf{A} \times \mathbf{P}|^+ = (\mathbf{A}^+ \times \mathbf{P}^+) \cup (\mathbf{A}^- \times \mathbf{P}^-),$$

etc. Also, any condition (a,p,u) with (a,p) in $|\mathbf{A} \times \mathbf{P}|^+$ and u in $\mathbf{U}^+$ must yield a value of the dependent variable which is greater than the constant value for conditions in $|\mathbf{A} \times \mathbf{P}|^0 \times \mathbf{U}$ or $\mathbf{A} \times \mathbf{P} \times \mathbf{U}^0$, etc., and the latter constant (the natural zero on the rescaled dependent variable) must be greater in turn than any condition in $|\mathbf{A} \times \mathbf{P}|^+ \times \mathbf{U}^-$, etc.

Altogether, the partitioning of factors into positive, negative, and neutral subsets, if it exists, should be fairly obvious, because the complete reversals of ordering should be a striking phenomenon, unlikely to be missed when it occurs. In some instances (see Example 7), the partitioning can be carried out in advance of collecting quantitative data.

Generalized Cancellation

We start the diagnostic procedure with the test of sign dependence: if it fails, then all simple polynomials are rejected; if it holds, then we test additional properties. Following the testing procedure developed for the positive case, we turn now to the test of double cancellation. Like independence, double cancellation need not hold in the signed case. It is satisfied, however, within any *homogeneous* subset of the factorial structure, that is, a subset in which no multiplicative factor changes sign. Clearly, in the positive case the entire structure is homogeneous. To test the various cancellation conditions in the signed case, the structure is divided into homogeneous subsets according to the partition of its factors. Since any homogeneous subset containing a neutral multiplicative factor has a degenerate or a trivial order, only the positive and the negative sets are considered. The multiplicative rule induces eight such homogeneous subsets (e.g., $\mathbf{A}^+ \times \mathbf{P}^+ \times \mathbf{U}^+$, $\mathbf{A}^+ \times \mathbf{P}^+ \times \mathbf{U}^-$, $\mathbf{A}^- \times \mathbf{P}^- \times \mathbf{U}^-$, etc.); the distributive rule induces four (i.e., $|\mathbf{A} \times \mathbf{P}|^+ \times \mathbf{U}^+$, $|\mathbf{A} \times \mathbf{P}|^+ \times \mathbf{U}^-$, $|\mathbf{A} \times \mathbf{P}|^- \times \mathbf{U}^+$, $|\mathbf{A} \times \mathbf{P}|^- \times \mathbf{U}^-$); and the dual-distributive rule induces four (i.e., $\mathbf{A}^+ \times \mathbf{P}^+ \times \mathbf{U}$, $\mathbf{A}^+ \times \mathbf{P}^- \times \mathbf{U}$, $\mathbf{A}^- \times \mathbf{P}^+ \times \mathbf{U}$, $\mathbf{A}^- \times \mathbf{P}^- \times \mathbf{U}$). Once these subsets are identified, the appropriate cancellation conditions can be applied to each of the subsets. For a particular composition rule to be valid, all the conditions it implies in the positive case should be satisfied within each homogeneous subset. Hence, double cancellation must be satisfied by all such subsets if any of the simple poynomials is valid.

The next step is to test the number of pairs of factors that satisfy joint sign dependence. If no pair of factors satisfies it, then all rules are rejected; if one or possibly two pairs of factors satisfy it, then both the distributive and the dual-distributive rules may be valid; if all three pairs of factors satisfy it, then the multiplicative rule is appropriate.

When exactly one pair of factors satisfies joint sign dependence, a very simple method for discriminating between the distributive and the dual-distributive case exists. (This is usually the case under these rules; see Footnote 9.) With no loss of generality, we can label these factors **A** and **P**. In the signed case, if the distributive rule holds, then either **U** is sign dependent, but not independent (of **A** and **P**), or **A** and **P** are jointly sign dependent, but not jointly independent (with respect to **U**), or both. On the other hand, if the dual-distributive rule holds, then both **U** is independent (of **A** and **P**), and **A** and **P** are jointly independent (of **U**). Thus, a comparison of the pattern of independence and sign-dependence relations is sufficient for choosing between the distributive and the dual-distributive rules in this case. This comparison, however, does not guarantee that any of them is satisfied. To test each of the rules, the appropriate cancellation condition should be tested in each of the homogeneous subsets. Finally, one has to verify that the necessary order relations between the homogeneous subsets are satisfied, that is, positive combinations should exceed neutral combinations, which, in turn, should exceed negative ones.

A flowchart summarizing the major steps in the diagnosis of simple polynomials in the signed case is presented in Figure 5. Note that the testing procedure for the signed case, outlined in Figure 5, closely parallels the testing procedure for the positive case, outlined in Figure 4. The main difference between the two procedures is in the replacement of independence by sign dependence, and the restriction of cancellation to homogeneous subsets.

An Empirical Example

An interesting application of the present methods, to the problem of diagnosing rules of composition for factors affecting perceived risk, can be found in the paper of Coombs and Huang (1970). Subjects were asked to order, with respect to risk (not preference!), games that consist of multiple independent plays of a simple monetary gamble. In such a gamble, one receives x if a fair coin comes up heads and loses y if the coin comes up tails. Each game of this type can be characterized as a triple (r,v,n), where $r = x + y$ is the range of the outcomes, $v = \frac{1}{2}(x - y)$ is the expected value of each play, and n is the number of independent repeated plays of the game (r,v).

An inspection of the effects of zero values suggests the distributive composition rule. First, note that for $n = 0$ the game becomes trivial, and all trivial games are equally risky. Hence **N** must be a multiplicative factor with $N(0) = 0$. Similarly, if both r and v equal zero, the game again becomes trivial; yet, if v alone equals zero, the game may have different degrees of risk, depending on r. Consequently, **R** × **V** (but not **V**) is also a multiplier, and the distributive rule, therefore, is likely to be the only appropriate composition rule, of all the three-variable simple polynomials.

In the actual study, three levels of **R**, three levels of **V**, and two levels of **N** were employed in a factorial design. To diagnose the correct composition rule, independence (Equation 1) was first investigated. Actually, Coombs and Huang (1970) tested a more general form of Equation 1, but it is evident from their data that independence, too, is almost perfectly satisfied (i.e., with no more than a single violation) by a majority of the subjects (i.e., 16 out of 28). Double cancellation was tested only for subsets with a fixed value of **N**. Since only two levels (1 and 5) of **N** were used, double cancellation could be tested only for **R** × **V** planes. Practically no violations of this property were observed.

Following the steps outlined in Figure 4, the authors tested joint independence (Equation 2). Allowing no more than a single violation, it was found that in the risk orderings of the majority of the subjects, **R** and **V** were jointly independent (of **N**); **V** and **N** were jointly independent (of **R**); but **R** and **N** failed to be jointly independent (of **V**). The number of subjects (out of 28) that satisfied joint independence (with no more than a single violation) were, respectively, 22, 17, and 7. Finally, distributive cancellation was perfectly satisfied by 21 subjects, and the overall average number of violations was less than one. Thus, the analysis clearly supports the distributive rule that was initially favored by Coombs and Huang (1970).

Critical Diagnostic Tests

The ordinal properties introduced thus far were designed not merely to diagnose the appropriate composition rule, but also to provide rather thorough tests of its validity. If we assume that one of the rules is satisfied, then simpler and more efficient critical tests are available for discovering the cor-

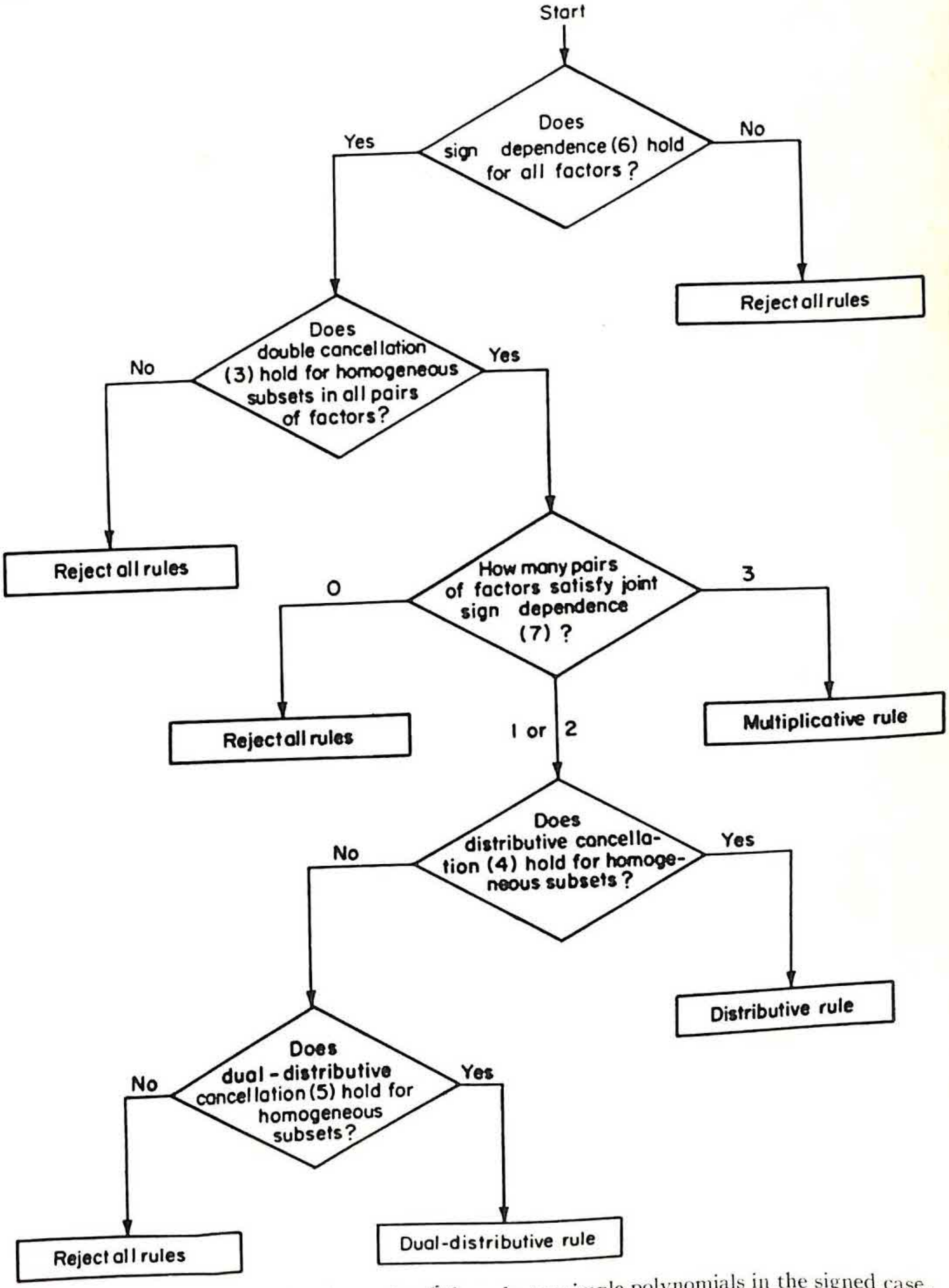

Start
Does sign dependence (6) hold for all factors?
Yes
No
Reject all rules
Does double cancellation (3) hold for homogeneous subsets in all pairs of factors?
No
Yes
Reject all rules
How many pairs of factors satisfy joint sign dependence (7)?
0
3
1 or 2
Reject all rules
Multiplicative rule
Does distributive cancellation (4) hold for homogeneous subsets?
No
Yes
Distributive rule
Does dual-distributive cancellation (5) hold for homogeneous subsets?
No
Yes
Reject all rules
Dual-distributive rule

FIG. 5. Flowchart for the diagnosis of three-factor simple polynomials in the signed case.

rect rule. In the first place, the signed properties of the variables usually determine the appropriate rule. For example, suppose that one of the simple polynomials is valid and that the ranges of the measurement scales of all three factors include zero. An element a_0 of **A** is said to be an *effective zero* for **P** if (a_0,u) induces equality over **P** for every u in **U**, that is, if $(a_0,p,u) = (a_0,q,u)$ for all p, q in **P** and u in **U**. In the light of this definition, it is readily seen that in this case, the number of factors containing an effective zero determines the appropriate composition rule. If there are three, the rule is multiplicative; if there are two, the rule is dual distributive; if there is one, the rule is distributive; and if there are none, the rule is additive. Further discussion of the diagnostic problem in the signed case can be found in Krantz et al. (1971, Ch. 7).

The development of efficient diagnostic procedures is not limited to the signed case. To demonstrate, we show how to design a critical test to choose among simple polynomials in the positive case. For convenience, we label the odd factor, if one exists, **U**. The candidate for the odd factor may be identified a priori, or it can be determined through the test of joint independence. Select treatment conditions so that the effects of a, p, and u are, respectively, larger than the effects of b, q, and v, and such that $(a,p,v) = (b,p,u)$. Now, compare (a,q,v) with (b,q,u). If the additive (or, equivalently, the multiplicative) rule holds, then

$$A(a) + P(p) + U(v) = \mathrm{A}(b) + P(p) + U(u);$$

adding $P(q) - P(p)$ to both sides yields

$$\mathrm{A}(a) + P(q) + U(v) = A(b) + P(q) + U(u),$$

which, in turn, implies $(a,q,v) = (b,q,u)$. (Note that this conclusion follows at once from joint independence.) On the other hand, if the distributive rule holds, then the equality $(a,p,v) = (b,p,u)$ implies

$$[A(a) + P(p)]\, U(v) = [A(b) + P(p)]\, U(u),$$

or

$$\frac{A(a) + P(p)}{A(b) + P(p)} = \frac{U(u)}{U(v)}.$$

Since the levels have been selected so that $A(a) > A(b)$, and $P(p) > P(q)$, and all values are assumed to be positive, it follows that

$$\frac{A(a) + P(q)}{A(b) + P(q)} > \frac{U(u)}{U(v)},$$

hence

$$[A(a) + P(q)]\, U(v) > [A(b) + P(q)]\, U(u),$$

which implies $(a,q,v) > (b,q,u)$. Finally, if the dual-distributive rule holds, then $(a,p,v) = (b,p,u)$ implies

$$A(a)\, P(p) + U(v) = A(b)\, P(p) + U(u),$$

or

$$P(p)[A(a) - A(b)] = U(u) - U(v);$$

and since, by hypothesis, $A(a) > A(b)$, $P(p) > P(q)$, $U(u) > U(v)$, and the values are all positive, it follows that

$$P(q)\,[A(a) - A(b)] < U(u) - U(v),$$

hence

$$A(a)\, P(q) + U(v) < A(b)\, P(q) + U(u),$$

which implies $(a,q,v) < (b,q,u)$. Under the above hypotheses, therefore, the additive, the distributive, and the dual-distributive rules can be readily diagnosed depending *only* on whether (a,q,v) is, respectively, equal to, greater than, or less than (b,q,u).

DISCUSSION

The procedures developed in the previous section do not exhaust the available methods for the diagnosis and testing of composition rules. Moreover, it can be shown that it is *impossible*, in general, to provide a (finite) list of necessary properties that are also sufficient, for any composition rule (see Scott & Suppes, 1958; Tversky, 1967a). Nevertheless, from an empirical viewpoint, the proposed independence and cancellation conditions are practically sufficient for the validity of the various rules, because the instances where these conditions are satisfied while the rule is invalid are quite rare. Furthermore, as was already mentioned, the proposed conditions are actually sufficient to establish the validity of the rules in question, if the factorial structure under study is sufficiently rich (i.e., dense), or regular (i.e., equally spaced). Contrary to some commonly held views that ordinal information is too "weak" for measurement, modern research has shown, in many forms and contexts, that when sufficiently many ordinal observations are available, not only the theory, but even the numerical scales are essentially determined. For rigorous proofs of sufficiency, including discussions of constructive scaling procedures, the reader is referred to Krantz et al. (1971). In the present discussion, we only comment briefly on scaling and on the problem of fallible data, and then discuss the merits of the axiomatic versus the numerical approach to the composition problem.

Scaling

Although the present paper is confined to the applications of conjoint-measurement theory to diagnosis and testing, it is instructive to examine how the theory is employed to construct scales of measurement. Suppose the additive model holds. Since the "origin" of all three scales is arbitrary, we can select any a_0 in $\mathbf{A}$, p_0 in $\mathbf{P}$, and u_0 in $\mathbf{U}$ and set $A(a_0) = P(p_0) = U(u_0) = 0$. Furthermore, since one unit is also arbitrary, we can select any a_1 in $\mathbf{A}$, with $(a_1,p_0,u_0) > (a_0,p_0,u_0)$, and set $A(a_1) = 1$. Next, we look for some p_1 in $\mathbf{P}$ and some u_1 in $\mathbf{U}$ that satisfy the equation $(a_1,p_0,u_0) = (a_0,p_1,u_0) = (a_0,p_0,u_1)$. Under the additive rule, therefore, $A(a_1) = P(p_1) = U(u_1) = 1$. Once these values are determined, we can continue the search for, say a_2 in $\mathbf{A}$ that satisfies the equation $(a_2,p_0,u_0) = (a_0,p_1,u_1)$ whence $A(a_2) = 2 = P(p_1) + U(u_1)$ since, in this notation, the scale value of any treatment combination is the sum of its indices. If the factorial structure under study is rich enough so that elements satisfying the desired equations can always be found, then the above process yields measurement scales by marking off equal units along each of the factors. By selecting an appropriately small unit, the scale value of any treatment combination can be obtained. (This method is analogous to the familiar process of physical measurement, where length, for example is measured in terms of the number of smaller units (e.g., centimeters) that are needed to "match" the length of any given object.) If, however, the data are not sufficiently rich to permit the above construction, then the scale values can only be determined approximately in terms of upper and lower bounds. Essentially the same constructive method is applicable to the other composition rules, though the existence of both additive and multiplicative factors complicates the procedure somewhat.

Obviously, the proposed measurement can be carried through only when the composition rule is known. For example, the very ordinal properties (joint independence) that are use to test the additive rule are required to ensure that the above construction is consistent. For example, the element a_2 of $\mathbf{A}$ can be defined as the solution of the equation $(a_2,p_0,u_0) = (a_1,p_1,u_0)$, or it can be defined as the solution to the equation $(a_2,p_0,u_1) = (a_1,p_1,u_1)$, which follows from the first equation by joint independence. If, however, joint independence does not hold, there is no reason to expect that the same element of $\mathbf{A}$, that is, a_2, will satisfy both equations. The necessary diagnostic properties, therefore, play a major role in the construction of the measurement scales. Furthermore, they can be tested in the actual scaling process by checking that the values otained by solving the various equations are, indeed, consistent.

Fallible Data

A major problem encountered in testing ordinal properties is the presence of error. If the data are infallible, then the necessary properties can be routinely checked and the test yields an unequivocal conclusion. Typically, however, the data are regarded as fallible in the sense that they reflect random fluctuations or sampling errors, in which case the testing problem is considerably more complicated. Independence, for example, may be violated by the ordering estimated from the data, merely as a consequence of sampling error. Any adequate evaluation of the validity of some composition rule, therefore, should take the error problem into account.

Two general approaches to this problem are available. First, we may collect sufficiently many observations to ensure the stability of the observed order and thereby reduce the probability of an erroneous conclusion to some acceptable low value. Although it is always desirable to obtain a more stable order, the above approach may be quite costly. Alternatively, we may interpret the ordinal properties as statistical hypotheses, and test them accordingly. For example, independence may be reformulated probabilistically as $Pr[(a,p,u) \geq (b,p,u)] \geq \frac{1}{2}$ if and only if $Pr[(a,q,v) \geq (b,q,v)] \geq \frac{1}{2}$, where Pr can be defined in terms of the proportion of times (or subjects) for which the inequality holds. For pairs of treatment combinations that are statistically independent, the above hypothesis can be tested using standard statistical tests such as a likelihood ratio test. If the treatment combinations cannot be treated as statistically independent, then more complicated statistical methods have to be employed.

Axiomatic versus Numerical Approach

The physical measurement of fundamental attributes such as mass, length, or time is based on certain qualitative laws (i.e., axioms) that lead to the *construction* of numerical scales on the basis of qualitative, or ordinal, observations. In contrast, most of the work (e.g., Luce, 1959; Stevens, 1957; Thurstone, 1927) on the measurement of psychological attributes such as loudness, intelligence, or subjective value has been based on some *given* numerical scales that are either produced by subjects (e.g., rating scales, magnitude estimates), or that describe, in physical units, some aspects of their behavior (e.g., choice probabilities, test scores). Similarly, most of the work on the composition problem in psychology (e.g., Anderson, 1970; Goldberg, 1968; Shepard, 1964; Thurstone & Jones, 1957) has been based on the assumption that the numerical values of the dependent variable can be regarded as interval-scale measures of the relevant psychological attribute.

Like the classical (extensive) theory of physical measurement, the theory of conjoint measurement is based on the analysis of qualitative laws. Unlike the physical theory, however, that is concerned with the laws that govern single attributes, the theory of conjoint measurement is concerned with the qualitative laws that govern the composition of several attributes. Inherent in the conjoint-measurement approach is the view that the importance of measurement (as distinct from other usages of numerical indices) lies in the qualitative laws on which it is based, and that the usefulness of numerical scales derives from the empirical lawfulness that they reflect. Accordingly, the measurement is regarded as a consequence, or a by-product, of the validity of some qualitative laws, and major emphasis is placed on the study of these laws. This view has not been shared by all. Some (e.g., Adams, 1966) proposed that the importance of numerical scales in science far transcends the axioms that characterize them, while others (e.g., Anderson, 1970) argued that it is preferable to study composition rules in terms of numerical values than in terms of qualitative properties.

In discussing the merits of the numerical and the axiomatic approaches to the study of composition rules, we submit that they should be regarded as complementary rather than competing: The two methods of analysis are likely to highlight different aspects of the data, and both analyses may prove useful in various contexts. With respect to the numerical approach, it should be noted that in some situations the numerical values of the dependent variable obey the proposed composition rule, to a satisfactory degree of approximation (e.g., Anderson, 1970; Gulliksen, 1956; Hoffman, 1960; Rimoldi, 1956; Spence & Guilford, 1933). In such instances, the dependent variable can be regarded as an adequate measure of the relevant psychological attribute, and the measurement problem is greatly simplified, because standard statistical procedures (e.g., linear regression, analysis of variance) can be used not only to test the validity of the proposed composition rule, but also to obtain estimates of the contributions of the various treatment conditions. It should be added, however, that in order to demonstrate the validity of some proposed composition rule, the statistical test must have adequate power. Otherwise, it may fail to reject a composition rule because of noisy data. Unfortunately, this important factor is all too often neglected.

If the numerical values of the dependent variable do not decompose according to the proposed composition rule, it may be *either* because the rule is invalid, *or* because the numerical scale is inadequate. At this point, an ordinal analysis is required to determine which of the possibilities obtains. The analysis may consist of testing the ordinal properties (i.e., axioms) that are necessary for the composition rule under study, along the lines described earlier in this paper.

An alternative to testing ordinal properties is to search for an appropriate monotonic transformation of the dependent variable that best satisfies the proposed rule, and then evaluate the correspondence between the proposed rule and the rescaled data. The desired monotonic transformation may be selected on the basis of an informal inspection of the data, or it can be constructed in a systematic fashion, such as by using an iterative computer program to solve a system of simultaneous inequalities dictated by the data. The availability of high-speed computer programs facilitates the applicability of this method, yet the problems of finding the most appropriate transformation and evaluating its goodness of fit do not have simple satisfactory solutions. These problems have been investigated particularly in the case of the (simple) additive rule, where some useful scaling methods and statistical testing procedures have been developed, for example, Box and Cox (1964) and Kruskal (1965). Like other scaling techniques, these methods always provide numerical estimates of the scale values, regardless of whether the rule in question is valid or not. Hence, if one is primarily concerned with scaling, rather than with diagnosis, the numerical methods are recommendable.

On the other hand, if one is primarily interested in psychological theory, rather than in scaling, the axiomatic approach has much to offer. It usually provides not only an explicit and precise formulation of the theory, but often a deeper and a more penetrating analysis of the structure under study. The assumption that the orders of the effects of drive, habit strength, and incentive are all jointly independent, for example, is psychologically more meaningful, and perhaps more plausible, than the assumption that response strength is a multiplicative function of drive, habit strength, and incentive. Similarly, the qualitative assumptions of expected-utility theory (e.g., transitivity, the extended sure-thing principle) provide a much more satisfactory formulation of decision making under uncertainty than the assertion that people behave as if they assign utilities and subjective probabilities to outcomes and events, respectively, multiply the utility of each outcome by its subjective probability of occurrence, and then choose the alternative for which the sum of these products is maximal. Although the axiomatic and the numerical formulations of the multiplicative or the expected-utility rules are equivalent, the former suggest a more plausible (nonnumerical) psychological mechanism. In this respect, it explicates the proposed numerical rule in terms of more fundamental, and often more transparent, principles.

The axiomatic analysis of the composition problem isolates the essential ordinal properties which underly the various composition rules. These properties, therefore, can be used to test the validity of the various rules in a measurement-free manner. Furthermore, they can be used to design critical tests for choosing among alternative rules, as demonstrated in the previous sections. The major practical advantage of the direct test of axioms over the general evaluation of numerical correspondence is that, in general, it provides a more powerful diagnostic test between alternative composition rules. The multiplicative and the distributive rules, for example, are so close to each other, in the positive case, that with fallible data it is extremely difficult to distinguish between them on the basis of any overall measure of goodness of fit. A comparison of the two rules in terms of joint independence, on the other hand, is likely to provide a much sharper and more diagnostic test, because it focuses on a critical distinguishing property. Furthermore, the test of critical properties is typically a more efficient method than an overall test of goodness of fit, in order to discover which aspects of a theory, if any, are at fault. An evaluation of goodness of fit of a numerical model can indicate that the theory is violated, but unlike the test of specific axioms, it is not very likely to locate the source of the discrepancy. The failures of joint independence in the color space (Krantz, in preparation), or transitivity in choice behavior (Tversky, 1969), or of the extended sure-thing principle (MacCrimmon, 1968) in decision making under uncertainty are all examples of significant empirical findings that were revealed by the direct test of critical axioms, and that were practically impossible to detect from global analyses of goodness of fit of the corresponding numerical models.

In summary, the conjoint-measurement approach, exemplified in this paper, offers a new methodology for the study of composition rules, with potential for deeper theoretical analysis and sharper empirical tests.

REFERENCES

Adams, E. W. On the nature and purpose of measurement. *Synthese*, 1966, **16**, 125–169.

Adams, E. W., & Fagot, R. F. A model of riskless choice. *Behavioral Science*, 1959, **4**, 1–10.

Anderson, L. R., & Fishbein, M. Prediction of attitude from the number, strength, and evaluative aspect of beliefs about the attitude object: A comparison of summation and congruity theories. *Journal of Personality and Social Psychology*, 1965, **2**, 437–443.

Anderson, N. H. Functional measurement and psychophysical judgment. *Psychological Review*, 1970, **77**, 153–170.

Atkinson, J. W. Motivational determinants of risk-taking behavior. *Psychological Review*, 1957, **64**, 359–372.

Box, G. E. P., & Cox, D. R. An analysis of transformations. *Journal of the Royal Statistical Society* (Series B), 1964, **26**, 211–252.

Brown, R. *Social psychology.* New York: The Free Press, 1965.

Cliff, N. Adverbs as multipliers. *Psychological Review*, 1959, **66**, 27–44.

Coombs, C. H., & Huang, L. Polynomial psychophysics of risk. *Journal of Mathematical Psychology*, 1970, **7**, 317–338.

Davidson, D., Suppes, P., & Siegel, S. *Decision-making: An experimental approach.* Stanford, Calif.: Stanford University Press, 1957.

Debreu, G. Topological methods in cardinal utility theory. In K. J. Arrow, S. Karlin, & P. Suppes (Eds.), *Mathematical methods in the social sciences, 1959.* (Stanford Symposium) Stanford, Calif.: Stanford University Press, 1960.

Goldberg, L. Simple models or simple processes? Some research on clinical judgments. *American Psychologist*, 1968, **23**, 483–496.

Gollob, H. F. Impression formation and word combination in sentences. *Journal of Personality and Social Psychology*, 1968, **10**, 341–353.

Gulliksen, H. Measurement of subjective values. *Psychometrika*, 1956, **21**, 229–244.

Hammond, K. R. Probabilistic functioning and the clinical method. *Psychological Review*, 1955, **62**, 255–262.

Heider, F. Attitudes and cognitive organization. *Journal of Psychology*, 1946, **21**, 107–112.

Hoffman, P. J. The paramorphic representation of clinical judgment. *Psychological Bulletin*, 1960, **57**, 116–131.

Hölder, O. Die Axiome der Quantität und die Lehre vom Mass. *Berichte über die Verhandlungen der Königlich Sächsischen Gesellschaft der Wissenschaften zu Leipzig, Mathematisch-Physische Classe*, 1901, **53**, 1–64.

Hull, C. L. *A behavior system.* New Haven: Yale University Press, 1952.

Hurvich, L. M., & Jameson, D. Some quantitative aspects of an opponent-colors theory. II. Brightness, saturation, and hue in normal and dichromatic vision. *Journal of the Optical Society of America*, 1955, **45**, 602–616.

Krantz, D. H. A survey of measurement theory. In G. B. Dantzig & A. F. Veinott, Jr. (Eds.), *Mathematics of the decision sciences, part 2.* Providence, R. I.: American Mathematical Society, 1968.

Krantz, D. H., Luce, R. D., Suppes, P., & Tversky, A. *Foundations of measurement.* New York: Academic Press, 1971, in press. 2 vols. (Volume 1, in press; Volume 2, in preparation)

Kruskal, J. B. Analysis of factorial experiments by estimating monotone transformations of the data. *Journal of the Royal Statistical Society* (Series B), 1965, **27**, 251–263.

Luce, R. D. *Individual choice behavior.* New York: Wiley, 1959.

Luce, R. D., & Krantz, D. H. Conditional expected utility. *Econometrica*, 1971, **39**, in press.

Luce, R. D., & Tukey, J. W. Simultaneous conjoint measurement: A new type of fundamental measurement. *Journal of Mathematical Psychology*, 1964, **1**, 1–27.

MACCRIMMON, K. R. Descriptive and normative implications of the decision theory postulates. In K. Borch (Ed.), *Conference on risk and uncertainty.* London: Macmillan, 1968.

PAVLIK, W. B., & REYNOLDS, W. F. Effects of deprivation schedule and reward magnitude on acquisition and extinction performance. *Journal of Comparative and Physiological Psychology,* 1963, **56**, 452–455.

RIMOLDI, H. J. A. Prediction of scale values for combined stimuli. *British Journal of Statistical Psychology,* 1956, **9**, 29–40.

ROSENBERG, M. J. Cognitive structure and attitudinal affect. *Journal of Abnormal and Social Psychology,* 1956, **53**, 367–372.

ROSKIES, R. A measurement axiomatization for an essentially multiplicative representation of two factors. *Journal of Mathematical Psychology,* 1965, **2**, 266–276.

SAVAGE, L. J. *The foundations of statistics.* New York: Wiley, 1954.

SCOTT, D., & SUPPES, P. Foundational aspects of theories of measurement. *Journal of Symbolic Logic,* 1958, **23**, 113–128.

SEWARD, J. P., SHEA, R. A., & DAVENPORT, R. H. Further evidence for the interaction of drive and reward. *American Journal of Psychology,* 1960, **73**, 370–379.

SHEPARD, R. N. On subjective optimum selection among multiattribute alternatives. In M. V. Shelly, II, & G. L. Bryan (Eds.), *Human judgments and optimality.* New York: Wiley, 1964.

SPENCE, K. W. *Behavior theory and conditioning.* New Haven: Yale University Press, 1956.

SPENCE, W., & GUILFORD, J. P. The affective value of combinations of odors. *American Journal of Psychology,* 1933, **45**, 495–501.

STEVENS, S. S. On the psychophysical law. *Psychological Review,* 1957, **64**, 153–181.

THURSTONE, L. L. A law of comparative judgment. *Psychological Review,* 1927, **34**, 273–286.

THURSTONE, L. L., & JONES, L. V. The rational origin for measuring subjective values. *Journal of the American Statistical Association,* 1957, **52**, 458–471.

TVERSKY, A. A general theory of polynomial conjoint measurement. *Journal of Mathematical Psychology,* 1967, **4**, 1–20. (a)

TVERSKY, A. Additivity, utility, and subjective probability. *Journal of Mathematical Psychology,* 1967, **4**, 175–201. (b)

TVERSKY, A. Intransitivity of preferences. *Psychological Review,* 1969, **76**, 31–48.

TVERSKY, A., & KRANTZ, D. H. Similarity of schematic faces: A test of interdimensional additivity. *Perception and Psychophysics,* 1969, **5**, 124–128.

VON NEUMANN, J., & MORGENSTERN, O. *Theory of games and economic behavior* (3rd ed.) Princeton: Princeton University Press, 1951.

(Received August 27, 1970)

Psychological Review
1971, Vol. 78, No. 2, 170

LETTERS TO THE EDITOR

PREPAREDNESS AND STRUCTURE

Seligman (Seligman, M. E. P. On the generality of the laws of learning. *Psychological Review,* 1970, 77, 406–418) presents convincing arguments for the position that the laws of learning are not general, but rather vary with the organism's preparedness for a particular association. Herein, he succeeds in his stated goal of bringing animal learning concepts into the mainstream of psychological interest. Associative preparedness is similar to the human skill learning concept of *s-r compatibility* (Fitts, P. M. Skill learning. In A. W. Melton (Ed.), *Categories of human learning.* New York: Academic Press, 1964). But more can be done.

Seligman ties preparedness to evolutionary history—*inherited* biological structure. Associability must depend on structure, but *personal* as well as evolutionary history is involved. Encounters between the organism and his environment determine the direction of development of inherited structures (e.g., Held, R., & Hein, A. Movement-produced stimulation in the development of visually guided behavior. *Journal of Comparative and Physiological Psychology,* 1963, **56**, 872–876). In humans, the dependence of associability on learned structures is now recognized (Mandler, G. Verbal learning. In T. Newcombe (Ed.), *New directions in psychology: III.* New York: Holt, Rinehart & Winston, 1967). Underwood and Schultz (Underwood, B. J., & Schultz, R. W. *Meaningfulness and verbal learning.* Philadelphia: Lippincott, 1960) point out that a major factor in paired-associate learning is the meaningfulness of the relation between stimulus and response. Meaningfulness is a *learned* relation.

The challenge, then, is to provide an empirically testable account of the development of structure. Piaget's theory is an example. Furthermore, the nature of specific structures could be used as the basis for predicting associability. Using Seligman's examples: Can Garcia's conditioned illness data and Thorndike's scratching data both depend on a psychobiological structure involving sharp differentiation between events outside, on the periphery, and inside the organism? If so, what kinds of experiences common to most mammals contribute to the development of this structure?

The advantage of this approach is that it provides an alternative between extremes: tremendously specific inherited preparednesses for all possible associations and the oversimplified one-kind-of-learning view which Seligman justly criticizes.

(Received November 24, 1970)

Charlotte Doyle
Department of Psychology
Sarah Lawrence College
Bronxville, New York 10708

DEVELOPMENTAL PSYCHOLOGY

Editor: Boyd R. McCandless

Developmental Psychology contains articles which represent the broad range of growth and development and their major associated variables. Chronological age as well as sex, socioeconomic status and efforts of physical growth variables (including time of maturation and body build) are all considered relevant development variables.

Although the adolescent and the aging population appear to deserve particular attention at this time, *Developmental Psychology* indeed covers the entire span. Cross-species articles and articles concerning developmental research with retardates are also included.

The articles in *Developmental Psychology* are not limited to any one approach such as "experiments versus studies" or "cognition versus motivation".

Bimonthly. Two volumes per year.
Subscription Price: $20.00 per year.

AMERICAN PSYCHOLOGICAL ASSOCIATION
1200 Seventeenth Street, N. W.
Washington, D. C. 20036

9 NOV 1971

Announcing a new book —

The Perception of Language

Edited by **D. L. Horton,** *University of Maryland*, and **James J. Jenkins,** *University of Minnesota*

February, 1971/256 pages/6x9″/$8.95/#9250

A publication in the MERRILL INTERNATIONAL PSYCHOLOGY SERIES edited by Arthur W. Melton and Donald B. Lindsley.

This book describes the proceedings of a conference held at the University of Pittsburgh on January 10-11, 1968. The conference brought together twenty psychologists and psycholinguists to present their particular research interests and to attempt to find communalities of thinking through discussion of "The Perception of Language."

The contributors include Harry Levin, Paul A. Kolers, Eleanor E. Maccoby, David McNeill, James G. Martin, Herbert H. Clark, Emerson Foulke, David B. Orr, Jerry A. Fodor, John B. Carroll, W. S. Condon, W. D. Ogston, Eric H. Lenneberg, Donald Shankweiler, Herbert Rubin, Katherine S. Harris, Kenneth N. Stevens, Malcolm S. Preston, and James Deese.

For a copy of *The Perception of Language*, please send your check for $8.95 and we will send it to you postpaid. You may return the book for full refund within 30 days if not satisfied. Please address your request to **BOYD LANE:**

CHARLES E. MERRILL PUBLISHING COMPANY
A Bell & Howell Company
1300 Alum Creek Drive
Columbus, Ohio 43216

Pocket 2
Issue 60/15 Ins
Librarian
[illegible]

PSYCHOLOGICAL REVIEW

GEORGE MANDLER, EDITOR
JEAN M. MANDLER, ASSOCIATE EDITOR
UNIVERSITY OF CALIFORNIA, SAN DIEGO

Vol. 78, No. 6 **November 1971**

Articles

Specific Hungers and Poison Avoidance as Adaptive Specializations of Learning
Paul Rozin and James W. Kalat 459

Adjectives, Comparatives, and Syllogisms
Janellen Huttenlocher and E. Tory Higgins 487

More about "Adjectives, Comparatives, and Syllogisms": A Reply to Huttenlocher and Higgins
Herbert H. Clark 505

Judgment and Response in Magnitude Estimation
John Ross and Vincent Di Lollo 515

Metacontrast and Lateral Inhibition . . **Bruce Bridgeman** 528

Temporal Auditory Acuity . . . . **David M. Green** 540

Theoretical Notes

Randomization and the Shape Function Model of Learning: A Reply to Wiesen . . . . **D. M. Warburton** 552

Ebbinghaus' Derived-List Experiments Reconsidered
Wayne Shebilske and Sheldon M. Ebenholtz 553

Letters to the Editor 556

This is the last issue of Volume 78.
Title page and Index for the volume appear herein.

Published Bimonthly
by The American Psychological Association, Inc.

Editorial Consultants

Norman H. Anderson
J. W. Atkinson
Nathan Azrin
Samuel Barondes
J. E. Birren
M. E. Bitterman
Herbert H. Clark
Peter J. Dean
David Elkind
William K. Estes
C. R. Gallistel
John Garcia
Rochel Gelman
Stephen E. Glickman
Lewis R. Goldberg
Stephen Handel
Jerry Higgins
J. Huttenlocher
E. Klima
David H. Krantz
John I. Lacey
G. David Lange
David Messick
Lance A. Miller
John Morton
Paul A. Obrist
Irwin Pollack
David Premack
Floyd Ratliff
George S. Reynolds
Julian B. Rotter
Paul Rozin
K. Warner Schaie
Wilbur Schramm
Norman J. Slamecka
Richard L. Solomon
J. E. R. Staddon
Lowell H. Storms
Robert Teghtsoonian
Warren Torgerson
Ledyard R. Tucker
Allan R. Wagner
D. Zeaman

(A large group of psychologists has kindly agreed to assist in the editing of manuscripts submitted to the *Psychological Review*. In each issue the Editor will list in the *Review* those editorial consultants who have helped him process manuscripts for that issue, so that their efforts may be gratefully acknowledged.)

The *Psychological Review* is devoted to articles of theoretical significance to any area of scientific endeavor in psychology. Except for occasional articles, manuscripts exceeding 7,500 words will be discouraged. Ordinarily, manuscripts which consist primarily of original reports of research or of reviews of the literature should be submitted to other journals. Methodological and statistical articles are not appropriate for the *Psychological Review*. See the editorials in the *Review* (1965, **72**, 1–2, 1971, **78**, 1–2) *for further details.*

Beginning with the 1971 issue, Letters to the Editor will be published on an experimental basis. Letters not exceeding 250 words on topics of theoretical interest should be submitted in duplicate.

Manuscripts must be accompanied by an abstract of 100 to 120 words typed on a separate sheet of paper. The abstract should conform to the style of *Psychological Abstracts*. Detailed instructions for preparation of the abstracts appeared in the *American Psychologist* (1961, 16, 833), or may be obtained from the APA Central Office.

All manuscripts must be submitted in triplicate. In the reference list, give journal titles in full; do not abbreviate. Figures should be submitted as glossy prints; duplicate figures may be photographic or pencil-drawn copies. Authors should retain a copy of the manuscript in case there is loss in the mail.

Because of the large number of manuscripts submitted, there is an inevitable publication lag of several months. Authors may avoid this delay if they are prepared to pay the costs of publishing their own articles; the appearance of articles by other contributors is not thereby delayed.

Manuscripts should be addressed to the Editor, George Mandler, Department of Experimental Psychology, South Parks Road, Oxford OX1 3PS, England.

Copyrighted material. The following policies govern reprinting or other forms of reproduction of materials copyrighted by the American Psychological Association: Written permission must be obtained for copying or reprinting any tables, figures, or text of more than 500 words in length. Persons or organizations who obtain permission to reproduce any such copyrighted materials must include on the first page of the reproduced materials the exact copyright notice which appeared on the Association's original publication. There will be a charge of $10.00 per page for such permission except when the requester is the author of the article to be reprinted, or when the article is to be reproduced in a limited number of copies solely for instructional purposes.

Helen Orr
Managing Editor

Lorraine Biros
Technical Editor

Henry S. Salmon
Advertising and Business Manager

Anne Redman
Subscription Manager

Published Bimonthly by the
AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
Prince and Lemon Streets, Lancaster, Pa.
1200 Seventeenth St., N. W., Washington, D. C. 20036

$10.00 volume

$2.00 issue

Second-class postage paid at Lancaster, Pa.

Send all communications, including address changes, to 1200 Seventeenth St., N. W., Washington, D. C. 20036. Address changes must arrive by the tenth of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

457
12/12

PSYCHOLOGICAL REVIEW

GEORGE MANDLER, EDITOR
JEAN M. MANDLER, ASSOCIATE EDITOR
UNIVERSITY OF CALIFORNIA, SAN DIEGO

VOLUME 78, 1971

HELEN ORR
Managing Editor

LORRAINE BIROS
Technical Editor

HENRY S. SALMON
Advertising and Business Manager

ANNE REDMAN
Subscription Manager

Published Bimonthly
by The American Psychological Association, Inc.

Editorial Consultants

Robert P. Abelson
Bernard Agronoff
Truett Allison
Anne Anastasi
Norman H. Anderson
Elliot Aronson
A. John Arrowood
J. W. Atkinson
Fred Attneave
Nathan Azrin
Alan D. Baddeley
Albert Bandura
Horace Barlow
Samuel Barondes
A. T. Beck
Jacob Beck
Daryl Bem
Daniel E. Berlyne
Ira H. Bernstein
Thomas G. Bever
R. L. Birdwhistell
J. E. Birren
M. E. Bitterman
Donald S. Blough
Gerald S. Blum
Robert C. Bolles
Lyle E. Bourne, Jr.
Gordon H. Bower
Martin D. Braine
Dana Bramel
Jerome Bruner
Theodore H. Bullock
Donn Byrne
Donald T. Campbell
J. Merrill Carlsmith
Dorwin Cartwright
A. Charles Catania
Herbert H. Clark
Barry E. Collins
Stewart Cook
Clyde H. Coombs
Stanley Coopersmith
T. N. Cornsweet
Tom Cornsweet
C. G. Costello
Nicholas B. Cottrell
Phebe Cramer
Robyn Dawes
Peter J. Dean
Clinton DeSoto
A. J. Deutsch
J. A. Deutsch
Bruce Earhard
M. Earhard
Howard Egeth
Paul Ekman
David Elkind
Seymour Epstein
Robert P. Erickson
William K. Estes
E. J. Fantino
Norman T. Feather
Martin E. Fishbein
Hans Furth
Robert M. Gagne
C. R. Gallistel
Leo Ganz
John Garcia
Wendell R. Garner
Rochel Gelman
Kenneth Gergen
Michael T. Ghiselin
James J. Gibson
Henry Gleitman
Stephen E. Glickman
Lewis R. Goldberg
Harry Gollob
Robert Gottsdanker
James Greeno
Lee W. Gregg
G. Robert Grice
S. P. Grossman
Ralph N. Haber
G. S. Halford
Stephen Handel
C. S. Harris
Jerry Higgins
Samuel Himmelfarb
James Hinrichs
Julian Hochberg
Paul Holland
Edwin Hollander
W. K. Honig
Janellen Huttenlocher
Francis W. Irwin
James R. Ison
Irving L. Janis
H. M. Jenkins
Lawrence Karlin
William Kessen
Charles Kiesler
Gregory A. Kimble
Walter Kintsch
E. Klima
Paul A. Kolers
David H. Krantz
John I. Lacey
Peter J. Lang
G. David Lange
Marvin Levine
A. M. Liberman
Frank A. Logan
Perry London
Frederic Lord
Elijah Lovejoy
R. Duncan Luce
Dominic Massaro
David C. McClelland
David McNeill
Ronald Melzack
David Messick
Samuel Messick
Lance A. Miller
Norman Miller
Allan F. Mirsky
Walter Mischel
John Morton
Bennett B. Murdock, Jr.
Paul H. Mussen
Ulric Neisser
Slater Newman
Raymond Nickerson
Clyde E. Noble
D. A. Norman
M. Frank Norman
Paul A. Obrist
Alan Paivio
Dean Peabody
Carl Pfaffman
Irwin Pollack
E. C. Poulton
David Premack
Karl H. Pribram
Floyd Ratliff
W. F. Prokasy
Frank Restle
George S. Reynolds
Bernard Rimland
Irvin Rock
Eliot H. Rodnick
Richard Rose
Julian B. Rotter
Paul Rozin
Richard Rudner
David E. Rumelhart
Irwin G. Sarason
K. Warner Schaie
Wilbur Schramm
Paul F. Secord
T. Shallice
Roger N. Shepard
Norman J. Slamecka
Richard L. Solomon
George Sperling
Sally Sperling
J. E. R. Staddon
Saul Sternberg
S. S. Stevens
Lowell H. Storms
Patrick Suppes
Renato Tagiuri
Percy Tannenbaum
Robert Teghtsoonian
H. S. Terrace
John Theios
W. R. Thompson
Warren Torgerson
Anne Treisman
Ledyard R. Tucker
Endel Tulving
Amos Tversky
Paul Vitz
Allen R. Wagner
Lise Wallach
Michael A. Wallach
Wilse Webb
William Wiest
Jerry Wiggins
H. A. Witkin
Joachim Wohlwill
Robert S. Wyer
Robert B. Zajonc
David Zeaman
Philip Zimbardo
Joseph L. Zinnes

CONTENTS OF VOLUME 78

ARTICLES

Anderson, Norman H. **Integration Theory and Attitude Change** ... 171

Berch, Daniel B. **Comments on Biederman's "Continuity Theory Revisited: A Failure in a Basic Assumption"** ... 260

Bernbach, Harley. **Strength Theory and Confidence Ratings in Recall** ... 338

Boies, Stephen J. See Posner, Michael I.

Bridgeman, Bruce. **Metacontrast and Lateral Inhibition** ... 528

Clark, Herbert H. **More about "Adjectives, Comparatives, and Syllogisms": A Reply to Huttenlocher and Higgins** ... 505

Coltheart, Max. **Visual Feature-Analyzers and Aftereffects of Tilt and Curvature** ... 114

Costello, C. G. **Conditioned Avoidance Responses and Phobias: A Reply to Wolpe and to Powell and Lumia** ... 348

Di Lollo, Vincent. See Ross, John.

Dodwell, P. C. **On Perceptual Clarity** ... 275

Dreyer, Paul, and Renner, K. Edward. **Self-Punitive Behavior—Masochism or Confusion?** ... 333

Dunham, Philip J. **Punishment: Method and Theory** ... 58

Ebenholtz, Sheldon M. See Shebilske, Wayne.

Feather, N. T. **Organization and Discrepancy in Cognitive Structures** ... 355

Frederiksen, John R. **Statistical Decision Model for Auditory Word Recognition** ... 409

Green, David M. **Temporal Auditory Acuity** ... 540

Halwes, Terry, and Jenkins, James J. **Problem of Serial Order in Behavior Is Not Resolved by Context-Sensitive Associative Memory Models** ... 122

Higgins, E. Tory. See Janellen Huttenlocher.

Huttenlocher, Janellen, and Higgins, E. Tory. **Adjectives, Comparatives, and Syllogisms** ... 487

Jackson, Douglas N. **The Dynamics of Structured Personality Tests: 1971** ... 229

Jenkins, James J. See Hawles, Terry.

Kalat, James W. See Rozin, Paul.

Krantz, David H., and Tversky, Amos. **Conjoint-Measurement Analysis of Composition Rules in Psychology** ... 151

Levine, Marvin. **Hypothesis Theory and Nonlearning Despite Ideal S-R-Reinforcement Contingencies** ... 130

Lumia, Augustus R. See Powell, Arnold.

Martin, Edwin. **Verbal Learning Theory and Independent Retrieval Phenomena** ... 314

Miller, Franklin G. See Weiss, Robert Frank.

Miller, Lance A. **Hypothesis Analysis of Conjunctive Concept-Learning Situations** ... 262

Moore, Dewey J., and Shiek, David A. **Toward a Theory of Early Infantile Autism** ... 451

Posner, Michael I., and Boies, Stephen J. **Components of Attention** ... 391

Postman, Leo. **Organization and Interference** ... 290

Powell, Arnold, and Lumia, Augustus R. **Avoidance Conditioning and Behavior Therapies: A Reply to Costello** ... 344

Renner, K. Edward. See Dreyer, Paul.

Ross, John, and Di Lollo, Vincent. **Judgment and Response in Magnitude Estimation** ... 515

Rozin, Paul, and Kalat, James W. **Specific Hungers and Poison Avoidance as Adaptive Specializations of Learning** ... 458

Shebilske, Wayne, and Ebenholtz, Sheldon M. **Ebbinghaus' Derived-List Experiments Reconsidered** ... 552

Shiek, David A. See Moore, Dewey J.

Simmelhag, Virginia L. See Staddon, J. E. R.

Staddon, J. E. R., and Simmelhag, Virginia L. **The "Superstition" Experiment: A Reexamination of Its Implications for the Principles of Adaptive Behavior** ... 3

Stevens, S. S. **Issues in Psychophysical Measurement** ... 426

Teghtsoonian, Robert. **On the Exponents in Stevens' Law and the Constant in Ekman's Law** ... 71

Thistlethwaite, Donald L. See Wellens, A. Rodney.

Thomas, Hoben. **Discrepancy Hypotheses: Methodological and Theoretical Considerations** ... 249

Todd, Thomas C. See Vitz, Paul C.
Toman, Walter. The Duplication Theorem of Social Relationships as Tested in the General Population 380
Treisman, Michel. On the Word Frequency Effect: Comments on the Papers by J. Catlin and L. H. Nakatani 420
Tversky, Amos. See Krantz, David H.
Vanderwolf, C. H. Limbic-Diencephalic Mechanisms of Voluntary Movement 83
Vitz, Paul C., and Todd, Thomas C. A Model of the Perception of Simple Geometric Figures 207
Warburton, D. M. Randomization and the Shape Function Model of Learning: A Reply to Wiesen 555
Weiss, Robert Frank, and Miller, Franklin G. The Drive Theory of Social Facilitation 44
Weitzman, Bernard. A Reply to Wolpe 352
Wellens, A. Rodney, and Thistlethwaite, Donald L. An Analysis of Two Quantitative Theories of Cognitive Balance 141
Wiesen, Joel P. Comment on "General Shape Function Model of Learning with Applications in Psychobiology" 272
Wolford, George. Function of Distinct Associations for Paired-Associate Performance 303
Wolpe, Joseph. The Behavioristic Conception of Neurosis: A Reply to Two Critics 341

LETTERS

Anderson, Norman H. An Exchange on Functional and Conjoint Measurement 457
Auerbach, Carl. Interdependence of Stevens' Exponents and Discriminabilty Measures 556
Bernbach, Harley A. Recognition Memory and Signal Detection Analyses 274
Bethlehem, Douglas W. On Acquiescence and Social Desirability 81
Brandt, Lewis W. Subjective Time and Decision Making 81
Doyle, Charlotte. Preparedness and Structure 170
Hirschberg, Morton A. See Kendler, Howard H.
Kendler, Howard H., Hirschberg, Morton A., and Wolford, George. Spence's Prediction about Reversal-Shift Behavior 354
Krantz, David H., and Tversky, Amos. An Exchange on Functional and Conjoint Measurement 457
Steinheiser, Frederick H., Jr. A Nonmediational Ideomotor Theory 354
Tversky, Amos. See Krantz, David H.
Wolford, George. See Kendler, Howard H.

PSYCHOLOGICAL REVIEW

VOL. 78, NO. 6 NOVEMBER 1971

SPECIFIC HUNGERS AND POISON AVOIDANCE AS ADAPTIVE SPECIALIZATIONS OF LEARNING[1]

PAUL ROZIN[2] AND JAMES W. KALAT[3]

University of Pennsylvania

Learning and memory are considered within an adaptive-evolutionary framework. This viewpoint is illustrated by an analysis of the role of learning in thiamine specific hunger. Consideration of the demands the environment makes on the rat and the contingencies it faces in the natural environment, appreciation of the importance of the novelty–familiarity dimension for these animals, and the realization of two new principles of learning, permit a learning explanation of most specific hungers. The two new principles, "belongingness" and "long-delay learning" specifically meet the peculiar demands of learning in the feeding system. In conjunction with the importance of the novelty dimension, they are discussed in an attempt to develop the laws of taste-aversion learning. It is argued that the laws or mechanism of learning are adapted to deal with particular types of problems and can be fully understood only in a naturalistic context. The "laws" of learning in the feeding system need not be the same as those in other systems; manifestation of a learning capacity in one area of behavior does not imply that it will be accessible in other areas. This notion leads to speculations concerning the evolution and development of learning abilities and cognitive function. Full understanding of learning and memory involves explanation of their diversity as well as the extraction of common general principles.

The application of the basic principles of evolution and adaptation to learning and memory offers a hopeful means of conceptualizing and ordering the increasingly diverse data of these fields. Learning and memory, being the result of natura selection, should be expected to be best developed in situations and species where other solutions to the problems at hand are less adaptive. Furthermore, when learning or memory capacities are brought to bear on a particular type of problem or situation, it stands to reason that these capacities should be shaped by or adapted to the situation.

We propose to treat learning and memory as any other biological characteristic, subject to natural selection and therefore adapted to handle specific types of problems. Insofar as an important environmental problem (e.g., obtaining adequate foods) presents unique demands or contingencies, we would expect to find appropriate modifications of learning and memory abilities, closely articulated with one another and with the natural behavior

---

[1] This manuscript is an expanded version of an invited address by the first author to APA Division 6 in 1969. The preparation of this article and some of the research reported in it were supported by National Science Foundation Grant GB 8013. We would like to thank John Garcia, Henry and Lila Gleitman, Elisabeth Rozin, Martin Seligman, W. J. Smith, and Richard Solomon for their helpful comments and criticisms.

[2] Requests for reprints should be sent to Paul Rozin, Department of Psychology, University of Pennsylvania, Philadelphia, Pennsylvania 19104.

[3] Supported by National Institute of Health training grant. Now at Duke University.

© 1971 by the American Psychological Association, Inc.

of the organism. Biologically speaking, there is no reason to assume that there should exist an extensive set of generally applicable laws of learning, independent of the situation in which they are manifested. This is not to say that there might not be some general laws, at a minimum resulting from basic constraints and features of the operation of the nervous system, and perhaps reflecting general principles of causality in the physical world. However, if we look at learning within an adaptive-evolutionary framework, we should seek not only to uncover some of the greatest common denominators among the behaviors we study, but also to explore the plasticity of the mechanisms themselves, as they are shaped through selection to deal with particular types of problems. For many years, the leading ethologists (e.g., Lorenz, 1965; Tinbergen, 1951) have espoused a position consistent with this. They have emphasized the importance of considering learning within a naturalistic context; learning is viewed as being genetically programmed to occur at specific points in an ongoing behavior sequence. The thesis presented here is in harmony with the ethologists' position, but emphasizes differences in learning *mechanisms* themselves, as a function of the situations in which learning occurs.

In the face of recent evidence, coming most significantly from the work of Garcia and his colleagues (Garcia, Ervin, & Koelling, 1966; Garcia & Koelling, 1966), challenging two important generalities in learning, a framework for ordering the suggested diversity of laws of learning is desirable. We propose such a framework in terms of the notion of adaptive specializations. Consideration of learning from the adaptive point of view offers two advantages: (*a*) it provides a heuristic, for ordering and predicting the types of learning one will see in different situations, and (*b*) it provides a type of explanation of behavior, in that one aspect of understanding behavior can be considered to be explication of its adaptive value.

Adaptive considerations have proved useful in many areas of biology and are a significant part of our understanding of sensory psychology and physiology. Specializations in sensory function, as shown in the classic works of Lettvin, Maturana, McCulloch, and Pitts (1959) in the visual system of the frog, Capranica (1966) on the remarkable tuning of the frog auditory system to mating calls, and Roeder (1963) on detection of bat sonar by the moth auditory system, can be fully understood only in the context of adaptations to natural problems. The incredible diversity of the visual system (e. g., variations in the proportion of rods and cones in the retina as a function of species and ecology) can be ordered and comprehended in a phylogenetic adaptive framework, as shown in the classical work of Walls (1942). We would like to suggest that there are similar possibilities in the area of learning and memory.

In the major portion of this paper, we shall discuss specific hungers and poisoning in rats. We shall argue that some basic features of learning and memory as applied to food selection in the rat are strikingly different from features characterizing the rat's learning in more traditional laboratory situations, that these differences make sense in terms of evolutionary adaptation, and that an understanding of the role of learning and memory in food selection involves an elucidation of specifically adapted learning mechanisms and an integration of these with genetically determined behavior patterns.

In the remaining portion of the paper, we shall discuss whether the feeding system is a unique case or representative example, relate our position to somewhat similar theoretical positions, including ethology and the recent papers of Garcia (Garcia & Ervin, 1968; Garcia, McGowan, & Green, 1971), Revusky (Revusky, in press; Revusky & Garcia, 1970), Seligman (1970), Lockard (1971), and Shettleworth (1971), and discuss the implications of our position in the areas of comparative psychology of learning and human function.

SPECIFIC HUNGERS

Basic Phenomenon and Problems with a Learning Interpretation

The phenomena of specific hungers were beautifully demonstrated in the classical work of Richter (1943). The question raised by this work was, How is the animal (in particular, the rat) able, when deficient in a particular nutrient, to select those foods containing it? In the case of sodium, it is quite clear that sodium deficiency releases an *innate* preference for substances containing sodium (Nachman 1962; Richter, 1956; Stricker & Wilson, 1970). However, it is hard to believe that the rat comes equipped with prewired recognition systems for each of the many substances for which it can show a specific hunger. The alternative view holds that rats can learn about what foods "make them better" when they are ill and in this manner come to select the proper nutrients (Harris, Clay, Hargreaves, & Ward, 1933; Scott & Quint, 1946). Rats would, in their lifetime, learn preferences for the tastes of only those foods associated with recovery from the deficiencies they happen to have experienced. This notion has the distinct advantage of simplicity in that it accounts for all of the specific hungers (except sodium) with one basic mechanism.

There was, in fact, some evidence for a learning interpretation of specific hungers. Scott and Verney (1947) offered a distinctively flavored vitamin supplemented food and an unflavored deficient food to deficient rats. After a preference developed for flavored, enriched food, the flavor was switched to the deficient food. By and large, rats now preferred the flavored deficient food, suggesting a learning mechanism.

The problem with a learning interpretation, so far as psychologists were concerned, was the long delay of reinforcement. The recovery (reward) effects produced by an ingested nutrient would not occur until many minutes or hours after ingestion. This poses serious problems for a learning mechanism within the context of traditional learning theory. If specific hungers were in fact learned, then some new type of learning had to be involved. This served as the impetus for our investigation of specific hungers.

The first demonstration of vitamin B hunger was that of Harris et al. (1933), who found that given a choice of three foods, one supplemented with B vitamins, B vitamin deficient rats would quickly come to choose the correct food. For convenience, Rozin, Wells, and Mayer (1964) decided to study a more simplified situation. One vitamin, thiamine, was selected, since a clear specific hunger for thiamine had been reported (Richter, Holt, & Barelare, 1937; Scott & Quint, 1946). The most pronounced symptoms of thiamine deficiency in the rat are weight loss and anorexia. (More details about this and other aspects of thiamine hunger are available in a review by Rozin, 1967a). Following injection or ingestion of thiamine, deficient rats show clear signs of reversal of symptoms within a few hours, at most. The basic design consisted of raising weanling rats on a thiamine deficient diet for 21 days, at which point they showed clear signs of deficiency. At this point they were offered a choice between this deficient diet and the same diet supplemented with thiamine.

Thiamine deficient animals strongly preferred the thiamine enriched choice, while control animals who were not deficient, but otherwise treated identically, did not. A few rats with a choice between deficient and *highly enriched* diet showed no clear preference during the first few choice days, but in the meantime ingested great amounts of thiamine and showed marked recovery. Subsequently, these rats developed a clear preference for the thiamine rich diet. It seemed highly unlikely that a rat would show a preference for what it would have needed a few days ago without showing that same preference when the thiamine presumably had reinforcing value. This anomalous observation was confirmed in an experiment in which rats were made deficient in thiamine and then recovered from deficiency by injection of high amounts of thiamine while consuming the deficient diet. After a period of recovery, rats were offered a choice of thiamine riched and deficient diets. Preferences the thiamine enriched diets in rats that d been deficient emerged at all the re- very intervals studied (Rozin, 1965).

This result raises a second problem for a arning interpretation of thiamine hunger: ow could rats develop a learned prefer- nce for something that at the time the reference appeared, had no known rein- orcing effects? If the preference developed ecause of the initial reinforcing effects f the thiamine, why did it not appear ntil much later? We shall call this roblem *Preference after recovery*.

Careful consideration of the specific hungers situation and contingencies gave rise to additional problems that could not be easily solved within the traditional learning framework. Given that a rat might learn with the delays of reinforcement that seem to hold here, how would the recovery be specifically associated with a particular food? If the animal had eaten from a number of the choices available (or both choices in the two-choice situation), how could one of these choices, presumably the one containing thiamine, show a specific increase in preference? In other words, if an ingestion of two or more foods is followed by recovery, how does one of these foods acquire positive properties? We shall call this third problem, *Which food?*

A logical extension of this notion raises the question of how food or the eating of food becomes associated with recovery when many other potential candidates for association exist in the environment. After all, following ingestion of foods including the vitamin enriched food, and before recovery can ensue, the rat performs many acts, such as chewing, grooming, exploring, and sleeping, and receives many stimuli. By what principle does he elect to respond to food stimuli instead of other stimuli? We call this fourth problem, *Why food?*

We have, then, four problems in the application of traditional learning principles to the explanation of specific hungers:

1. Delay.
2. Preference after recovery.
3. Which food?
4. Why food?

Reinterpretation of Specific Hungers

The first step in the resolution of the problem as formulated came with the appreciation of the importance of novelty in the determination of food preferences in deficient rats (Rodgers & Rozin, 1966). Deficient rats show an immediate and almost complete preference for new foods, even when the new food is thiamine deficient and the old food has a thiamine supplement. In this case, the new-food preference reverses after a few days. The results *suggested* that the "specific hunger" for thiamine may be simply a reflection of the thiamine deficient rat's preference for new foods: If the new food contains thiamine, a learned preference could develop (mechanism still unknown).

Rodgers (1967a) succeeded in demonstrating quite conclusively that there is no specificity to thiamine specific hunger. Reasoning that the novelty response might overwhelm an existing tendency to prefer the vitamin, Rodgers offered deficient rats the choice of a deficient novel diet or the same novel diet supplemented with thiamine. The usual strong specific hunger did not appear. Furthermore, when the choice was between two different novel diets, one enriched, preferences for the enriched source did not develop rapidly. Finally, separate groups of thiamine deficient rats and pyridoxine deficient rats were offered their basal diet in two forms: one supplemented with thiamine, the other with pyridoxine. If there were any specificity, one would expect each group to show a preference for the vitamin that would produce recovery. There were no significant differences between the groups.

There seemed to be two types of explanations for the novelty effect. One is that the deficiency experience triggers an innately programmed "neophilia." A more attractive hypothesis is that the novelty effect comes about through learning. The paradigms described all pit the new food against the familiar food that has been associated with deficiency. The new-food preference could be a reflection of an aversion to the old food. Put more colloquially, we could ask whether preferences for new foods appear because the rat "loves the new stuff or hates the old stuff." The issue can be clarified by watching the behavior of deficient animals toward different foods (Rozin, 1967b). Rats housed in relatively large individual cages were observed during the first 15 minutes of an 8-hour food presentation each day. Their responses were classified into a number of categories, such as grooming, sleeping, eating, and chewing. The experimental rats were placed on a thiamine deficient diet. As deficiency developed, less time was spent eating during the first 15 minutes, and two striking behaviors, quite rare in normals, appeared. One was spillage of the food: the rat would approach the food cup, sniff at the food, and then paw at the food in a scooping motion, spilling it out of the cup, so that it fell through the wire mesh floor. (This spillage phenomenon in deficient rats was actually quite familiar, since we have always had some difficulty measuring the food intake of deficient rats on this account.) The other new behavior was what we call "redirected feeding." Following this initial investigation of the food cup, rats would occasionally move over to the wooden barrier separating the nest area from the rest of the cage and begin to chew it vigorously. They would also, occasionally, chew on the cage wires. These two behaviors suggested an aversion to the familiar diet. The redirected feeding suggests conflict between desire to eat and the aversiveness of the food offered. Spillage is often seen in normal rats with highly unpalatable foods, such as quinine adulterated diets.

When offered the old deficient food in a new container (metal instead of glass) in a new location, these deficient rats showed little potentiation of eating and continued to show the behavior described above. Apparently the vessel and its location were not controlling the aversion. However, when a completely new deficient food was offered in the familiar vessel and location, uninterrupted avid eating ensued, suggesting that the aversion was specific to the food. The deficient rats were subsequently allowed to recover on a new, vitamin enriched diet. Following one week of recovery, and after a 16-hour period of food deprivation, they were presented again with the familiar deficient food for the first time since the onset of recovery. These food deprived rats, showing no signs of deficiency at this time, responded to the familiar deficient food as they had before, with minimal ingestion and occasional spillage and redirected feeding. In this case, a "normal" hungry rat prefers staying hungry to eating its original diet which provides, in fact, perfectly adequate nutrition for him at the time, and is normally quite palatable. Preference of hunger (eating nothing) to ingestion in hungry rats and the similarity in the rat's behavior toward deficient and highly unpalatable (quinine adulterated) diets suggests strongly that we are dealing with an aversion to the familiar food.

The learned aversion interpretation places specific hungers in a new perspective. We can consider the deficient diet as a CS and the nausea or other ill effects produced by its ingestion as a UCS. Presumably, the classically conditioned "ill effects" lead to avoidance of the familiar food. The mechanism suggested for the specific hunger phenomenon, as seen in the standard two-diet choice, would then be that the rat learns an aversion to the familiar deficient food. Before the time of choice, he has already done a most significant part of his learning: He knows what not to eat. The initial preference for the new food follows. Its maintenance when the new food is enriched could be accounted for as an additional learned preference for the new food or as a failure to develop an aversion to it (see section on learned preference).

To the extent that specific aversions play a key role in specific hungers, there is an obvious parallel between specific hungers and poisoning. Both involve learned aversions; vitamin deficient diet is a slowly acting poison. The aversion experiment thus suggests that these two sets of phenomena are closely related. Furthermore, the same basic problems raised here (Delay, Which food? Why food?) arise with respect to poisoning. Since poisons are designed and synthesized by man, it is unreasonable to hold that rats show innate aversions to them. In comparing the literature on poisoning (see Barnett, 1963; Richter, 1953; Rźoska, 1953) to that on specific hungers, there is one apparent contrast. We have reported an increased *preference* for new foods in white rats, while the poisoning literature which focuses on wild rats strongly indicates the opposite: an exaggerated neo-*phobia* following poisoning. That is, wild rats, who show a much greater base-line tendency to avoid new objects or events than do white rats (e.g., see Galef, 1970), show a further exaggeration of this tendency, often to an extreme (Richter, 1953) following poisoning experiences. This disparity can be accounted for as a procedural difference, since the new-familiar choices we offered to our domestic rats were different from those usually offered poisoned wild rats. In particular, in the novelty experiments, which were done before we realized that specific aversions were involved, the rat was offered a choice between a familiar food associated with deficiency and a new food. Even if the white rat were somewhat neophobic, this might not appear since the alternative choice was a strongly aversive familiar diet.

In order to provide a meaningful comparison between poisoning and specific hungers, both sets of phenomena would have to be demonstrated under the same sets of conditions and in the same strain of rats. In an experiment meeting these criteria and comparing half-wild and domestic rats, a paradigm was employed that allowed fuller expression of the rat's neophobic or neophilic tendencies. Rats were raised on Diet A, prior to induction of aversive consequences. Deficiency or poisoning (or nothing in the case of controls) occurred in the presence of Diet B. In the final test, rats were offered these two diets and a completely new one, Diet C. Therefore, rats were faced with a choice among a familiar *safe* diet (A), a familiar aversive diet (B), and a completely new diet (C) (Rozin, 1968).

The single important result is that all rats suffering poisoning or deficiency showed an increased preference for the familiar safe food, that is, a *neophobia*. Half-wild rats showed a stronger neophobia following the aversive experience, but half-wild controls also showed a higher base-line neophobia. All experimental rats almost completely avoided the familiar aversive food (B), but ate some of the completely new food. There were no major differences between the specific hunger and poisoning groups. Therefore, it appears that we can consider specific hungers as a parallel to poisoning. Again, the behavior makes sense in an adaptive framework: following an unpleasant food-related experience, the rat tends to return to a known, *safe* food.

Resolution of the Inconsistencies with a Learning Interpretation

The specific aversion explanation of specific hungers, the realization of the importance of the familiarity-novelty dimension, and the appearance of two articles of major significance by Garcia and his colleagues enable us to resolve the four basic problems with a learning interpretation.

Preference after recovery. This can be accounted for as a retained aversion to the familiar food. Rats made deficient in thiamine and then recovered by injection for 12 hours or 5 days (while continuing to eat the same deficient diet) show a strong initial novel-food preference when offered a novel-familiar food choice, with thiamine *absent* from both choices (Rozin & Rodgers, 1967). In these experiments, recovery took place in the presence of the familiar food. The existence of a novelty effect suggests that the aversion had not fully extinguished by the time of testing. It is significant in this regard that in most of our experiments (some unpublished) the recovery effect showed some diminution as time of recovery increased. The recovery effect should be maximal when recovery is associated with a new food, so that the old diet aversion does not extinguish. Strong "recovery' effects are in fact seen under such circumstances in the aversion experiment, with a single choice (Rozin, 1967b) and in typical two-choice experiments with long recovery periods (Rozin, unpublished data).

Which food? In the "standard' two-choice situation, this problem is easy to deal with; the rat learns to avoid the deficient food when it is the only diet available. There is no Which food? problem here. Because of this learned aversion, rats show an immediate, virtually complete preference for the novel food, thus making it possible for them to learn about its consequences.

The testing situations described up to this point, with the partial exception of the testing environment in the aversion experiment, have been very limited, well defined, and unnatural. Rats normally live in a much larger area than was permitted to them in these studied, have an elaborate social life, and seldom face simple binary food choices. It is not unreasonable to assume that one or only a few foods might be available to the rat during the deficiency period in nature, since with a wide variety of foods available, a deficiency would be quite unlikely. In their classic work on B vitamin hungers, Harris et al. (1933) offered deficient rats a choice among a large variety of foods, with only one containing B vitamins in significant amounts. By and large, they found that with large numbers of choices (6–10 foods), rats were unable to select the vitamin enriched source, and they found it necessary to "educate" the rats by offering only the enriched food for a period of days in order to establish a maintained preference for this food. Examination of the day-to-day intake of these animals before the education period suggests a distinctly nonrandom food selection pattern: food intake seemed to be restricted to only a few of the many choices on any particular day. This seemed to be an adaptive way of unconfounding the situation and suggested an analysis of the meal patterns of deficient rats faced with the choice of a number of new foods (Rozin, 1969a). A thiamine supplement was placed in one of four diet choices for each rat. Feeding was restricted to eight hours a day, and the food intake from each cup was measured at hourly intervals. Four of the 10 deficient rats studied developed clear preferences for the enriched choice within a few days, and two other rats developed strong preferences for two of the four choices, where one of the preferred choices was enriched. Analysis of the meals indicates a characteristic pattern, both on a daily and hourly basis. Means, except for an initial daily sampling of many or all of the choices, tended to be restricted to one food. The rat's feeding pattern maximizes the possibility of associating each diet with its appropriate consequences, since meals tend to be isolated in time and to consist of a single food. The emergence of a strong preference for the enriched food is, in each case, preceded by a clearly defined meal of that food. Furthermore, no rat failed to develop a clear enriched food preference if it ate an isolated meal of at least 1 gram from the enriched food. The rats that failed to show adaptive preferences in the initial part of the experiment failed to sample significant amounts of the enriched food. Similar, though less well defined, sampling patterns are seen in normal rats. Normals mix meals more, but this may be a direct consequence of the fact that their meals are larger. The anorexia of deficiency may, in and of itself, exaggerate in an adaptive way the normal feeding pattern of the rat. It is clear that we have here another part of the answer to the Which food? problem.

Finally, social factors have become implicated in food selection in some significant recent experiments. Galef and Clark (in press) studied responses to poisoning in colonies of wild rats observed in the laboratory under seminaturalistic conditions. They offered a group of wild rats two foods, one of which was poisoned. After a short period of time, all rats avoided this food. The poison was removed, but the rats still completely avoided the base. A litter was born, and the behavior of the young toward these new foods was observed. The parents and young were fed for three hours a day and were constantly under observation during this period. During a two-week period (Days 14–28 of life in the pups), the pups came out to eat, but ate only the safe food (possibly because this was the only food being eaten by the parents). When the young were fully weaned and placed in a new cage, separated from their parents, they continued for a six-day period to eat only the safe diet and ignore the formerly poisoned diet. Further experiments (Galef 1971; Galef & Clark, in press) suggest a straightforward explanation for this effect. Rat pups tend to approach areas where adults are located and begin feeding there. In this way, they become familiar with the foods eaten by the adults and avoid alternative diets as a result of their neophobia. More recent work (Galef & Clark, 1971; Galef & Henderson, in press) suggests an additional mechanism: chemical cues from mother's milk seem to function to familiarize the young with certain characteristics of the mother's food and produce initial postweaning preferences.

Why food? Rats in the aversion experiment (Rozin, 1967b) did not show an aversion to the container or the location of the deficient food. While introduction of a new food dramatically increased ingestion, change to a new location and container did not. Apparently what the animal learned was specific to the food.

In an independent and much more compelling experiment, Garcia and Koelling (1966) provided evidence that there was a specific tendency for taste "CSs" to be associated with certain types of visceral "UCSs," while exteroceptive CSs such as light and sound would be preferentially associated with exteroceptive UCSs, such as shock. They paired light, sound, and taste simultaneously with either electric shock or poisoning in different groups of animals. The shocked animals developed an avoidance of the light and sound, but not of the taste. The poisoned animals, subjected to the same light-sound-taste pairing, avoided the taste and not the sound or light.

Delay. The reinterpretation of specific hungers, in itself, does not bring us much closer to solving this important problem. It does not appear to be possible to reinterpret what appeared to be long-delay learning in terms of short delays. However, a critical experiment by Garcia et al. (1966) demonstrates that long-delay learning can occur in this system. They induced an aversion to saccharin in rats by injections of apomorphine, a drug that presumably produces gastrointestinal upset. Aversions were produced when the interval between termination of drinking and injection of the drug was one-half hour or more. Only a few trials were necessary to establish a clear aversion.

A New Problem—Are There Learned Preferences as well as Learned Aversions?

Before we leave the problem of specific hungers, there is one new issue to resolve. Much of the specific hunger phenomenon can be explained in terms of aversion; Is there a positive side also?

Up to this point we have provided evidence for three categories of foods based on the animal's experience (Rozin, 1968). These are unexperienced or novel, familiar-safe, and familiar-aversive. The question is, Is there a fourth category, familiar-positive? That is, is the experiencing of positive consequences any different from the experiencing of neutral consequences when these experiences are contingent upon ingestion of a particular substance?

Much of the evidence which appears to demonstrate positive preferences can be reinterpreted in terms of learned aversions. For example, Harris et al. (1933) found that when only one of a large number of diets contained adequate thiamine, most deficient rats were not successful in selecting the correct diet. However, if the rats were "educated" by being offered *only* the adequate food for several days, they showed a preference for this food when the larger number of diets were again offered. This experiment does not distinguish between learned-preference and learned-aversion interpretations. Rats could be avoiding the nonthiamine diets by a combination of learned aversions and neophobia. That is, after "education" the rats have learned that the thiamine-containing diet is "safe," but it is not certain whether they have learned that it has any special "recovery from deficiency" properties which distinguish it from other safe diets, since we know that deficient rats prefer "old safe" foods to either old aversive or novel foods (Rozin, 1968). Similarly, in the sampling experiment (Rozin, 1969a), the fact that the rat eventually prefers the only enriched diet in a four-choice situation could mean that he had developed aversions to the other choices.

There are some recent experiments which present more serious challenges to a pure aversion model. Garcia, Ervin, Yorke, and Koelling (1967) repeatedly made rats thiamine deficient and produced recovery by thiamine injection. Just prior to thiamine injection, rats drank saccharin solution. At all other times, water was the only fluid available. When thiamine deficient, the rats showed an increase of saccharin intake over trials, both with respect to their own water intake and to the saccharin intake of similarly treated animals where thiamine injection was not contingent on saccharin ingestion. Campbell (1969) has also demonstrated that thiamine deficient rats show an increase in their absolute intake of a sucrose solution which has been associated with recovery from deficiency. Zahorik and Maier (1969) used Garcia et al.'s (1967) procedure but with a modified test and found that rats prefer the taste associated with recovery to both the taste associated with deficiency and a novel taste. Furthermore, this preference was apparent in both deficient and recovered rats.

These experiments, which certainly provide evidence for learned preferences, can nonetheless be explained in terms of the three basic categories of food. A more decisive experiment would have to show that rats prefer a "recovery" solution to an old "safe" solution, that is, one they drank without aversive consequences at a time when then were not vitamin deficient.

Revusky (1967; Revusky & Garcia, 1970) performed a series of experiments to demonstrate that food with clear positive consequences would be preferred to foods with relatively neutral consequences. In his simple design, rats were fed one nutrient solution while hungry and a different one when satiated. After five days of this training, a significant preference developed in a two-bottle test for the solution drunk while deprived. This result is interpreted in terms of the greater (delayed) reinforcing effect of the solution drunk during deprivation. Since both solutions were equally familiar and the rats drank the solution offered in satiated conditions voluntarily (so that an acquired aversion would be unlikely), this experiment meets the requirements for demonstrating a learned preference. The effect is clear, though not large by comparison to the aversion data (see Revusky & Garcia, 1970, for additional data).

It is noteworthy that the positive preference effects reported have been rather small by comparison with learned aversions, and difficult to obtain (Kalat & Rozin, unpublished; Revusky & Garcia, 1970). We cannot satisfactorily explain this asymmetry. Possibly the rat is better prepared to learn aversions because rapid learning there has particular survival value; that is, mistakes are very costly. This remains, for the moment, an intriguing problem, with possibly great implications for the regulation of food intake.

Other Specific Hungers

We have offered a theory of thiamine specific hunger. We believe that it holds for other learned specific hungers as well. Novel-food preferences, which imply a learned aversion mechanism, have been demonstrated in calcium, sodium, and magnesium deficient rats (Rodgers, 1967a) and in pyridoxine and riboflavin deficient rats (Rozin & Rodgers, 1967). The preference after recovery phenomenon appears in identical form in thiamine, riboflavin, and pyridoxine deficiency (Rozin & Rodgers, 1967).

It seems reasonable to assume that the anorexia characteristic of most vitamin deficiencies reflects, at least in part, a learned aversion to the deficient food. This is dramatically clear in the case of thiamine deficiency, where the anorexia symptom disappears precipitously when a new diet is offered. On the other hand, there is relatively little anorexia in vitamin D or vitamin A deficiency, and in both cases there has *not* been a clear demonstration of a specific hunger.

There is an impressive body of research on amino acid specific hungers and responses to amino acid imbalance (Harper, 1967). These "deficiencies" are characterized by anorexia and seem to fit into the scheme we have described. Recently, Rogers and Harper (1970) presented evidence for a positive preference for a solution that corrects an amino acid imbalance.

We cannot complete this reconsideration of specific hungers without mentioning one particularly serious shortcoming of all the mechanisms we have discussed. Although, in principle, they can account for most of the individual vitamin or mineral specific hungers, it is not clear how the classic "cafeteria" self-selection of Richter (1943) can be accounted for. Rats appear to self-select an extremely well balanced diet. Unless we assume that they develop incipient deficiencies of a variety of nutrients and learn aversions and preferences on the basis of these minimal symptons and their abatement, we have no explanation of this remarkable phenomenon. At this time, long-delay learning mechanisms appear inadequate to the task, since we cannot identify obvious candidates for unconditioned stimuli.[4]

WHY IS SODIUM HUNGER INNATE?

An alternative to the mechanism we have discussed is an almost completely genetically preprogrammed specific hunger. Sodium hunger seems to be such a case. Rats sodium deficient for the first time show a strong preference for sodium immediately upon tasting it (Nachman, 1962; Richter, 1936, 1956; Stricker & Wilson, 1970). They also show strong preferences for similar-tasting salts which will not actually alleviate their deficiency (Nachman, 1962), and they will work at a high rate in extinction at a lever which previously produced a salt solution which they drank while sodium replete (Krieckhaus & Wolf, 1968).

We do not know why sodium differs from other specific hungers, but at least three explanations are available. First, the great importance of sodium in body fluid balance, its common scarcity in the environment (witness salt licks and salt wars), and the large amount needed by the individual organism might place particular selection pressure in favor of a surefire sodium hunger. The various physiological adaptions directed toward regulation of body sodium are impressive. Second, it is conceivable, as Rodgers (1967b) has suggested, that sodium ingestion by sodium deficient organisms may have initial negative effects[5] associated with large-scale electrolyte shifts. There is not much evidence on this point. Sodium deficient animals do develop an aversion to sodium deficient diet (Rodgers, 1967a). However, *injections* of sodium into sodium deficient animals have not successfully produced preferences for foods ingested just prior to injection (Rodgers, 1967b). These possible initial negative events may prevent the operation of the learning mechanisms normally implicated in specific hungers. Third, sodium as a stimulus is well defined by the taste system. It would be easy for a genetic program to tie into the proper class of stimuli. (Conversely, one might argue that the fact that sodium recognition seems to be so fundamental to the taste system suggests again the great importance of sodium as a directive force in evolution.) Although a unique learning ability might have evolved to meet this particular problem, the actual genetically programmed solution seems preferable. However, important biological functions are often overdetermined; this may be one example. It has been shown that normal rats can learn the location of sodium sources in their environment and go to them promptly when a sodium need is induced (Krieckhaus & Wolf, 1968).

---

[4] Another remaining problem concerns the difficulty in demonstrating specific hungers with the nutrients in an aqueous solution (Appledorf & Tannenbaum, 1967; Rozin, Wells, & Mayer, 1964). This contrasts with Richter, Holt, and Barelare's (1937) dramatic demonstration of a specific hunger for thiamine in solution. We are inclined to believe that the difficulty frequently encountered with aqueous solutions derives from pure water being one of the choices available. Water's great familiarity might protect it from becoming aversive (Garcia & Koelling, 1967; Maier, Zahorik, & Albin, 1971; Nachman, 1970a). Palatability differences in the various choices solutions might also be involved Rogers & Harper, 1970).

[5] Scott, Verney, and Morissey (1950) noted that magnesium deficient rats selected magnesium deficient food in preference to magnesium enriched food. Rodgers (1967a) confirmed this "inverse" specific hunger and suggested that the initial ingestion of magnesium by magnesium deficient rats produces aversive consequences. Here, those consequences might be considered as "coming down" from the hyperirritability or "high" characteristic of magnesium deficiency.

But now we must turn the question around and ask, Why are all specific hungers not innate? One basic learning mechanism can provide an adaptive solution for almost all specific hungers (and also poisoning). Instead of programming specific innate recognitions of a variety of nutrients, the organism can rely on the fact that the malaise produced by most deficiencies (and possibly the consequent recovery on ingestion of the needed nutrient) is adequate to establish aversions (and preferences). There remains the fascinating question of whether long-delay learning originally evolved in response to proper selection of nutrients or the regulation of food (caloric) intake, or both.

Sodium hunger raises the general issue of the factors influencing the role learning will play in solving a particular problem. With alternative mechanisms available to deal with an environmental problem, the solution achieved by a species should be determined by factors of economy (cost) dictated by the actual situation and features of the development of the organism. We can see the interplay of these factors in the case of what Stricker (personal communication, 1971) has pointed out is another specific hunger: thirst for water. Morgan (1894) studied the development of drinking in young chicks and found that they apparently had no innate recognition of water. For example, they would run through water trays without drinking. At some point in the first few days of life, the bird would ingest water by pecking at a water drop or at an irregularity on the surface or in the bottom of the water pan. As soon as water entered the bird's mouth, it would adopt characteristic drinking movements, and after only one or two such experiences, would show a clear visual recognition of water. Hunt and Smith (1967) repeated and extended these results.

This system is highly predetermined: all components seem to be genetically programmed, except the recognition of the visual stimulus for water. Even a reasonably precise regulation of water intake is programmed, ready to come into operation as soon as water is experienced. Stricker and Sterritt (1967) have shown that on its very first drink, a chick's water intake is proportional to its deficit. We suggest that the visual recognition of water is learned because this is an efficient way of accomplishing the task at hand. The situation in the environment is such that a bird looking for food (defined as small irregularities) will certainly end up with water in his beak, and since he is prepared to learn this and has all of the other genetic equipment necessary to handle water ingestion and balance, he is "home free." We emphasize that this mechanism is successful because the bird's normal pecking and eating behavior (genetically programmed) invariably brings it into a position where water enters the beak. Hunt and Smith (1967) show that the first time a bird pecks at a dew drop (likely to be his first experience with water in the real world), it responds with a "feeding" peck, which is characteristically different from an approach to water. Only after the water enters the beak does drinking-type behavior appear, and subsequently, the peck at dew drops can be seen to be qualitatively different from the peck at food.

We are suggesting that there are two types of advantages (and therefore evolutionary pressures) to learned solutions to problems. The first, and most obvious, has to do with the variable environment. In the face of a changing and variable environment, it would be maladaptive to program specific modes of response, since these would often be inappropriate. Clearly, a plastic organism can do better in a plastic environment. (We expect that monophagous animals are less capable of learning about food than omnivorous ones.) The second factor predisposing to learning has to do with the "cost of preprogramming." This has two components. First, preprogramming of a particular behavior or stimulus configuration preempts a portion of the total genetic material, which could otherwise be used for some other purpose. Second, there may be greater costs associated with the evolution of genetically programmed solutions, insofar as built-in solutions involve more genetic reorganization.

Of course, in some cases, a genetically programmed recognition may be *replaced* by a learning mechanism. If an animal had the capability of learning something that was genetically programmed, otherwise favorable mutations in the genetic material responsible for this behavior could be selected for, even though the genetic basis for the behavior would be destroyed. Given a general learning capability, it could often be cheaper to allow the environment to supply the appropriate stimulus configurations, even if these will not vary much throughout life. We suggest that this "balance sheet" type of approach may be fruitfully applied to all specific hungers. In the case of the chick, we suggest that it is necessary to genetically specify those portions of the feeding and drinking system which are distinctively different (the ingestive responses and regulation). Since the chick's behavior leads to water in the beak *anyway*, because the preprogrammed food recognition overlaps with this aspect of drinking, a cheap way out can be found. The costs are minimal, and the genetic savings, significant.

SUMMARY

The reanalysis of specific hungers rests on three fundamental points: (*a*) The actual contingencies in the feeding situation are not what they were thought to be. (*b*) The novelty-familiarity dimension is of particular importance to the rat. (*c*) To account fully for the phenomenon, two new "principles" of learning were needed: belongingness and long delay. It is remarkable that each of Garcia et al.'s two "classic" experiments (Garcia & Koelling, 1966; Garcia et al., 1966) dealt directly with one of these problems at just the time that these two issues became *the* two problems in specific hungers.

In the context of the problems of specific hungers, it seemed clear that the basic principles demonstrated in these experiments must be essentially correct. Both principles, belongingness and long-delay learning, seem highly adapted to the properties of the feeding system. Tastes are in fact causally linked to gastrointestinal events, and there is a long inherent delay between the taste and its consequences. We suggest that specific learning mechanisms have evolved in response to specific problems. In the following section, we shall consider taste-aversion learning in greater detail, in order to demonstrate the extensive ways in which learning mechanisms may be modified to satisfy particular demands.

TASTE-AVERSION LEARNING: AN EXAMPLE OF ADAPTIVE SPECIALIZATION OF LEARNING

In this section we shall consider learning about food. In considering whether it is a "new type" of learning, we shall examine both the belongingness and delay principles in some detail, consider the importance of familiarity and novelty, and explore other possible differences between learning in the feeding system and more traditional types of learning.

Principles of Stimulus Selection

We can presume, from the material on specific hungers, that when faced with a bewildering array of stimuli as candidates for association with a gastrointestinal event, the rat has available principles with which to sort them out. One concerns his past experience with these stimuli, the novelty-familiarity dimension, and the other certain presumably built-in tendencies to associate certain categories of stimuli with certain relevant events (belongingness). We shall consider each in turn.

Novelty-Familiarity Dimension and Associability

We have argued that the novelty (or familiarity) of a stimulus is of particular importance for a rat (see Galef, 1970; Rozin, 1968). This distinction has particular significance in determining the magnitude of a learned aversion to a given taste. Revusky and Bedarf (1967) and Wittlin and Brookshire (1968) showed that rats learn aversions much more readily to novel than to familiar solutions, even when the familiar solution is drunk after the novel solution (and, of course, before poisoning). In these experiments, familiarization occurred over a period of days, but, in fact, a 20-minute exposure to a solution followed by neutral consequences will produce virtually the same effect, for such a solution is then quite resistant to association with poisoning (Kalat & Rozin[6]). This minimal (single) experience can have occurred three weeks before the poisoning without significant attenuation of the effect.

Belongingness

The "belongingness" (Garcia & Koelling, 1966) or "preparedness" (Seligman, 1970) or "stimulus relevance" (Capretta, 1961) phenomenon—that is, the tendency to associate tastes with aversive internal consequences as opposed to associating either element with anything else—seems eminently sensible from an adaptive point of view. Gastrointestinal and related internal events are, in fact, very likely to be initiated or influenced by substances eaten, and taste receptors, by virtue of their location, provide information about these same substances. It is essential for an omnivore to have the ability to learn rapidly which substances to eat and which to avoid. However, an equal ability to associate lights and sounds with gastrointestinal consequences would be far less adaptive; in fact, the common result would be "superstitious" learning. The belongingness phenomenon receives support not only from these adaptive arguments but also from neurological considerations. The gustatory receptors and the gut receptors are similarly classified as visceral sensory inputs and show close neurological relationships, specifically in the medulla and possibly in the hypothalamus.

It appears that under *some* circumstances, exteroceptive cues can become associated with poisoning (Garcia, Kimeldorf, & Hunt, 1961; Rozin, 1969b). Rather rapid learning of aversion to locations and shape of drinking vessel can be demonstrated if the UCS (apomorphine) is introduced *during* drinking from the appropriate vessel (Rozin, 1969b). On the other hand, even with a modest 30-minute CS–UCS interval and the use of a concatenation of nontaste cues including vessel shape, location, solution texture, and temperature, virtually no aversion appears after many trials (Rozin, 1969; but see Nachman, 1970a, for a demonstration that under proper circumstances oral temperature—a nontaste but interoceptive cue—alone may be an effective cue). The suggestion here is that at least part of the "weakness" of exteroceptive cues derives from a very rapid decay of their associability with time.

The belongingness principle in relation to taste-aversion learning is elaborated by Garcia and Ervin (1968) and Garcia et al. (1971). Further extensions of the principle to areas outside of feeding have been discussed by Seligman (1970) and Shettleworth (1971).[7]

---

[6] J. Kalat and P. Rozin. Learned safety as an explanation for taste-aversion delay of reinforcement gradients. In preparation.

[7] The principle of belongingness makes some predictions about what aversive stimuli should be employed in various types of aversion therapy. For example, it would be expected that nausea-producing drugs such as apomorphine should be more effective than electric shock in treating alcoholism. At the present time, the data are not clear on this point (Rachman & Teasdale, 1969). The issue is complicated because it may be the case that humans learn different things with electric shock or nausea as UCSs.

Salience. There is evidence for "intramodality" belongingness. Rats tend to associate poisoning with some novel solutions more than others (Kalat & Rozin, 1970). Rats drank one novel solution briefly, 15 minutes later drank a second novel solution, and another 15 minutes later were poisoned. The following day, the rats were offered both solutions simultaneously. Under these conditions, certain solutions, which we describe as highly salient, became more aversive than others. The salience of a solution proved to be a more potent predictor of amount of acquired aversion than temporal proximity to poisoning. It was found possible to rank novel solutions in a stable, transitive "salience" hierarchy, such that each solution proved more salient (associable with poisoning) than all solutions lower on the list. Evidence for a salience effect on the "positive" side (recovery from thiamine deficiency) has been reported (Campbell, 1969).

As yet it is not known what constitutes the defining characteristics of the salience dimension. It is probable that the "relative novelty" of these operationally novel solutions contribute to the effect. By this we mean that more salient solutions may be more different from previously experienced solutions. There may be additional factors associated with intrinsic properties of the solutions. Certain tastes (e.g., bitter) may have a particular tendency to be associated with aversive consequences.

A functional definition of belongingness. At this time we know that some category of CSs, including tastes, preferentially associates with some category of UCSs, including "gastrointestinal distress." We would like to define these categories more precisely. The unconditioned stimuli employed have been described variously as "poisoning," "nausea," and "gastrointestinal upset." But the category of effective stimuli may include stimuli not suggested by these terms. Ingested foods can certainly produce significant internal effects outside of the gastrointestinal system, and it is indeed possible that the primary action of some of the UCSs presently employed is outside of the gastrointestinal system. On the other hand, it is hard to understand, from an adaptive point of view, why pain in the lungs or heart, for example, should be selectively associated with taste. As a matter of fact, it is not known whether gastrointestinal *pain* is selectively associated with taste. Much more research is needed to better define the range of visceral sensations with which tastes have high associability. The extensive Russian literature on interoceptive conditioning (see Bykov, 1957) demonstrates that exteroceptive cues can be attached to visceral UCSs. Whether there is some belongingness operating here remains to be seen.

Yet another dimension of belongingness concerns the temporal parameters of the various CSs and UCSs. In the taste-visceral system, stimuli tend to have slow onsets and to last for long periods. Exteroceptive CSs and UCSs, in contrast, are characteristically brief and well defined in time. The importance of these dimensions and the visceral field for taste-aversion learning could both be determined by using experimentally controlled UCSs, such as electric shock to the stomach mucosa instead of the ill-defined poisoning procedures.

The category of effective CSs should differ from species to species. It was argued above that the specific ability to learn taste-poison associations is highly adaptive because foods are the usual cause of any aversive gastrointestinal stimulation. This argument assumes that the animal identifies its food by taste. However, some species use other modalities as well. In particular, there is reason to believe that birds put main emphasis on visual cues in the identification and selection of food. A number of experiments by Brower (1969) indicate that birds can readily learn aversions to the sight of food,[8] and experiments by Wilcoxon, Dragoin, and Kral (1971) indicate that Japanese quail learn poison-based aversions more readily to the color (or optical density) than to the taste of a solution. Rats and bobwhite quail were poisoned half an hour after drinking a solution which was either blue, sour, or both blue and sour. The rats subsequently showed considerable aversion to the sour water, but none to the blue (dense) water. The quail, however, showed an aversion to both; furthermore, the quail poisoned on blue-sour water generalized their aversion to blue water and not to sour water. To show that the aversion to blue water was not based on the taste of the blue coloring, the investigators showed that quail could form an aversion to plain water in a distinctively colored tube.

---

[8] One type of mimicry is based on the fact that many birds and some other species readily learn to avoid toxic or unpalatable prey. In these cases, a safe, palatable species evolves coloration similar to that of an unsafe or unpalatable one and thereby obtains some protection due to the predator's generalization of its learned aversion (Brower, 1969; Rothschild, 1967; Wickler, 1968).

We suggest that the critical dimension for poison-based aversion learning may not be "taste versus other modalities" but "eating-related cues versus other cues." This type of functional categorization of input is in harmony with Gibson's (1966) view of perceptual systems. Eating-related cues would be whatever type of cue—gustatory, visual, or otherwise—a particular species uses to identify food. Because of the intimate and relatively invariant relationship between taste receptors and food ingestion, and because of the neurological association between taste and visceral receptors, it is likely that taste cues would always be classified as eating-related so that taste-poisoning belongingness should be practically universal. However, in some species, other modalities may also be eating related.

Nontaste sensory modalities could include both eating-related and non-eating-related cues. For instance, it may be that birds can form poison-based aversions to the sight of potential prey, but not to other sights. Also, it is known that odors are less effective than tastes in poison-based aversions in rats. Possibly odors are more effective when they emanate from a food source or if they are experienced simultaneously with a taste. It would be interesting to investigate the effective cues for poison-based aversion learning in other species in which food recognition is known to occur partly via nongustatory cues. For instance, frogs have specific visual cells which respond maximally to flying insects and similar stimuli (Lettvin et al., 1959). The above analysis suggests that stimuli which excite these "bug detectors' might be more easily associated with poison than other types of visual stimuli.

LONG-DELAY LEARNING

Until recently it appeared that learning of both classical and operant types was almost impossible with delays of reinforcement exceeding a few seconds. The importance of close temporal contiguity has been demonstrated in a considerable variety of experimental paradigims, and apparent exceptions seemed to depend on secondary reinforcement (Kimble, 1961). There were cogent theoretical and adaptive reasons for assuming the universal importance of close temporal contiguity. Premonitions existed, however, of a possible problem with learning about food ingestion.

History

The first such reference is Pavlov (1927). Using morphine as a UCS, Pavlov noted signs of nausea and profuse salivation in response to the touch of the experimenter, which preceded injection. Needless to say, the UCS (i.e., nausea, etc.) did not occur until many minutes after the morphine was administered. Pavlov evidently regarded the effect as a species of conditioning, thus implicitly accepting the possibility of learning over long delays:

> This experiment provides a clue to the well-known fact that dogs will eat meat the first time it is offered them, after removal of their parathyroids, or after an Eck fistula and tying of the portal vein, but on all subsequent occasions will refuse it. Evidently in these cases the appearance and smell of meat produce of themselves a reaction identical with that produced through direct pathological action in the absence of the parathyroids or the portal circulation, by those toxic substances resulting from digestion of the meat [p. 36].

A similar recognition of the phenomenon of long-delay learning is present in the biological literature on poisoning, regulation of food intake, and specific hungers. Thus, Harris et al. (1933), Scott and Verney (1947), Rźoska (1953), and LeMagnen (1969) explicitly implicated learning mechanisms; to biologists without any particular commitments to psychological theory, this explanation seemed perfectly plausible.

One way of reconciling these data to traditional S–R learning theory was to assume that animals associated a food with the consequences of a previous meal of the same food, thus achieving temporal contiguity (Rozin et al., 1964; Smith & Capretta, 1956). But rats need only a single exposure to a toxic food to learn to avoid it (Rźoska, 1953); thus some other explanation is necessary.

The discovery that long CS–UCS intervals are possible in learning about the consequences of foods occurred gradually. Garcia et al. (1961) in their early work found that rats learned food aversions on the basis of simultaneous exposure to X rays. Although it was known that X rays had their main effect with a considerable delay ("radiation sickness"), it was assumed that some immediate effect of X rays was involved in the learning, and slight evidence was offered to support this position. McLaurin (1964) was the first to operationally manipulate CS–UCS intervals over a wide range in taste-aversion learning with long delays of reinforcement. However, a methodological flaw precluded meaningful interpretation of this result; McLaurin tested for aversion immediately after exposure to X rays, and it was later demonstrated (Smith & Schaeffer, 1967) that the rats were learning aversions to the test solution during the test itself. That is, the drinking of the solution during testing was simultaneous with the development of aversive consequences of the X rays, and temporal contiguity was achieved. Garcia et al. (1966) avoided this problem by giving the test for saccharin aversion three days after exposure to apomorphine. This experiment successfully demonstrated learning with long delays of reinforcement of the order of *one hour.* Smith and Roll (1967) found similar results using X rays and either saccharin (up to 12-hour delays) or sucrose (up to 6-hour delays). Replications by Revusky (1968) using sucrose CS and X-ray UCS and Rozin (1969b) (using saccharin or casein hydrolysate as CSs and apomorphine as UCS) confirmed this effect.

Hypotheses to Explain the Long Delay

Peripheral—the aftertaste theory. Because of the revolutionary nature of this finding, there was considerable interest in the possibility that the apparent absence of temporal contiguity was illusory. Although the CS–UCS interval was ostensibly long, some peripheral trace of the CS might remain throughout the interval in the form of an aftertaste, regurgitation, or a high concentration in the blood. It could be this trace that was involved in the conditioning. However, a fair amount of data are now available to oppose this explanation. The main import of the data is that taste-aversion learning is possible under conditions that should greatly minimize any aftertaste.

First, taste-aversion learning to sucrose and saccharin solutions occurs on a single trial with maximum delays of poisoning of about 7 and 12 hours, respectively (Revusky, 1968; Smith & Roll, 1967). It is very doubtful that an aftertaste or perceptible change in blood concentration persists for such long periods. It is even less plausible to suggest that enough sucrose or saccharin remains in the stomach at this time to be tasted by regurgitation. Actually the regurgitation model never had much applicability to rats anyway, since rats rarely if ever vomit. A second argument against the aftertaste and related models is the fact that rats can learn aversions to a .05% HCl solution with a one-hour delay of reinforcement (Garcia, Green, & McGowan, 1969). A litmus paper test showed no measurable amount of HCl on the tongue two minutes after the animal stopped drinking. Thus, the likelihood of a conventional peripheral aftertaste an hour later is minimal. Furthermore, the solution was less concentrated than the HCl already in the stomach. Consequently, this experiment is peculiarly effective in eliminating regurgitation, stomach-tasting, and blood-tasting models. A third line of evidence is Rozin's (1969b) demonstration that rats can learn an aversion to a specific concentration of a solution as opposed to other concentrations of the same solution. The animal learned the aversion just as easily to the lower as to the higher concentration. Presumably after the 30-minute delay used in this experiment, the aftertaste of a more concentrated solution should taste more like a less concentrated solution of the same substance. Similarly, the blood or stomach concentration of the substance at the time of poisoning should not be uniquely related to a particular concentration of original solution by a function known independently by the rats. A fourth line of evidence is Nachman's (1970a) demonstration that rats can learn an aversion to a particular temperature of water; an aftertaste of a temperature is difficult to imagine. Fifth, rats can learn an aversion to a solution even if one or more solutions is drunk between the first solution and poisoning (Kalat & Rozin, 1970, 1971a; Revusky & Bedarf, 1967). The intervening solutions would surely minimize any aftertaste of the first. Finally, it has been argued (Revusky & Garcia, 1970) that even if the aftertaste model were correct, it would be difficult to reconcile the taste-aversion learning results to conventional learning theories. Even if there is an aftertaste, it would have been present and declining in intensity for such a long period that there would be no precedent for expecting learning to occur—let alone in one trial.

It may be argued that none of these arguments completely demolishes an aftertaste theory, and that certain other tests could be conducted, for example, attempting to produce aversions to solutions intubated intragastrically. Some recent evidence suggests that if such aversions occur at all, they are much smaller than taste-mediated aversions (Smith, 1970). However, even if demonstrable, intragastric aversions would not solve the delay problem, as the HCl experiments (Garcia et al., 1969), concentration aversions (Rozin, 1969b), and solution temperature aversions (Nachman, 1970a) are no easier to explain in terms of contiguity with intragastric stimuli.

Bradley and Mistretta (1971) have demonstrated the development of aversions to intravenously introduced solutions (saccharin) in rats, using X rays as the UCS. A circulating "slug" of high-concentration saccharin stimulates taste receptors in the tongue. This interesting experiment provides *another* mechanism of learned aversions, but it cannot be the only mechanism involved in orally mediated aversions. The blood concentrations used in these experiments are much higher than any which would occur naturally or in the long-delay experiments, and the same experiments discussed above as raising problems for an intragastric-tasting mechanism would be equally troublesome for a "blood-tasting mechanism."

At present there is no evidence in favor of an aftertaste theory and a considerable body of evidence against it. Establishment of an alternative theory seems a more appropriate strategy than accumulation of still further evidence against aftertaste interpretations. Of course, from the point of view of this paper, long-delay learning is exactly what should be expected in this situation, and the central-mediation alternative appears quite acceptable.

Central mechanisms—interference. Revusky (in press; see also Revusky & Garcia, 1970) proposes that the maximum CS–UCS interval in all types of learning depends not on time per se, but on the number of other potential CSs that intervene between the CS being tested and the UCS. That is, the animal tends to associate the UCS with the most recent potential CS, or perhaps the last several such stimuli. As the CS–UCS interval expands, the probability increases that the organism will perceive other sights, sounds, etc., and consequently the probability decreases that the animal will associate the UCS with the CS in question. In taste-aversion learning, the argument holds, the range of potential CSs is much more restricted. Only tastes have a substantial tendency to be associated with aversive gastrointestinal stimulation, and typically the test solution is the most recent taste the animal experienced prior to poisoning. Since no other potential CSs are present, there is nothing to interfere with learning over long delays. In taste aversion as in other learning, the animal tends to associate a UCS with the last potential CS; the only difference is that in taste-aversion learning, the potential CSs are fewer and less frequent. We would like to point out that the taste modality differs from the commonly studied exteroceptive modalities, in that virtually all stimulation comes from contacts initiated by behavior. Rats do not taste unless they approach something and introduce it into their mouth.

This theory is very attractive because it proposes that the differences between taste-aversion learning and other types of learning may all be derived from the general principle of belongingness, without postulating an independent difference in the delay of reinforcement function.

Nevertheless, the theory, if taken as the sole of primary explanation of delay, faces two serious problems. It predicts that in the absence of interfering taste cues, there should be little or no decrement in learning as the CS–UCS interval is increased. This is not the case. Garcia et al. (1966), Nachman, (1970a), Revusky (1968), and Kalat and Rozin (1971a) have all found an orderly decrease of aversion with increasing CS–UCS interval. Furthermore, all experiments have found a maximal interval, varying from about 2 to 12 hours, beyond which no learned aversion can be demonstrated. Kalat and Rozin (1971a) deprived rats of both food and water during the CS–UCS interval, and still observed an orderly decrease in learned aversions with increasing CS–UCS interval.

Not only is there a decrement in learning in the absence of explicit interference, but the addition of explicit interfering cues does not markedly reduce learning. The consumption of two or even three salient novel solutions during a 30-minute delay between sucrose consumption and poison does not eliminate the sucrose aversion (Kalat & Rozin, 1971a). The *maximal* interference effect we observed using three novel interfering solutions following sucrose, and one poisoning, left sucrose equally palatable with water, to which it is normally strongly preferred.

These experiments, then, suggest that retroactive interference can clearly weaken a potential association, but it is highly unlikely that the normal delay function can be largely accounted for in these terms: What could interfere with drinking of a test solution during six hours of no eating or drinking that would be *more* effective than three novel solutions? Proactive interference is an unlikely explanation. Effects from past taste experiences should be minimal, since most rats have experienced only highly familiar rat chow, water, and mother's milk.

Recent findings by Wilcoxon et al. (1971) pose further problems for Revusky's theory. Unlike rats, bobwhite quail poisoned 30 minutes after drinking unflavored water from a blue tube learn an aversion to drinking from that tube. Since the quail presumably saw a great many visual stimuli in the 30-minute delay, their ability to learn over long delays in this situation cannot be explained simply in terms of absence of interference.

Although it seems unlikely that interference represents the only basis for the CS–UCS delay gradient, it does appear to be a contributing factor. Since a rat normally experiences fewer tastes than other stimuli within a given period, the Revusky theory may account for part of the difference in the delay gradients between taste-aversion and other types of learning.

Slow decay of associability as an inherent property of the taste system. An alternative to the view that the delay-of-reinforcement gradient is a function of interference is the view that the delay gradient is an inherent property of the system itself; memory or associability decays much more slowly for taste than for other modalities. This leaves unanswered the question, What accounts for this decay?

One possibility is that the delay function represents the decay of short-term memory. According to this hypothesis, a stimulus is associable with other events only while it is held in short-term memory, and tastes remain in short-term memory for unusually long periods. (Although there is extensive data on short-term memory in humans, virtually none has been collected on taste, and it is conceivable that tastes do not compete with exteroceptive cues for space in short-term memory.)

One way of testing this hypothesis is by means of electroconvulsive shock, which has been assumed to eradicate short-term memories or to block their conversion to long-term memories. If taste associations must be made from short-term memory extended in time, it should be possible to demonstrate disruptive effects of electroconvulsive shock, presented within the CS–UCS interval but with longer delays following the CS than are ordinarily effective in disrupting other types of memory. In a very careful study, Nachman (1970b) found amnesic effects of electroconvulsive shock in some rats when electroconvulsive shock was presented immediately after 10 seconds of drinking but not after 30 seconds of drinking. He also found some amnesic effects from electroconvulsive shock presented 25 seconds after a 5-second drinking period. Kral (1970) also found small amnesic effects with electroconvulsive shock, using delays of 2 or 25 *minutes* following a 10-minute drinking period. As is usual with electroconvulsive shock experiments, the temporal parameters seem to vary widely from one experiment to another. However, in both cases, the effect of electroconvulsive shock is small; it impeded but did not prevent learning. And in both cases, the effective electroconvulsive shock times were within the range of times reported for electroconvulsive-shock-amnesic effects in other systems. Thus there is no evidence that the transfer of taste stimuli into long-term memory is unusually slow, or that tastes remain in short-term memory for long time.

Another possible mechanism for the dela function is that some central long-terr "trace" of the taste is decaying over time Unfortunately, this is a difficult hypothesi to test. One experiment (Rozin & Ree[9] at least puts certain constraints on th type of decay which is possible. Rat were anesthetized for 6–8 hours during th interval between consumption of a sucros solution and LiCl poisoning. These rats showed strong learned aversions at delays of poisoning considerably longer than those which are effective in the absence of anesthesia. Thus anesthesia seems to retard whatever process mediates the delay of reinforcement gradient. If this process is to be described as "decay," it is evidently an active rather than a passive type.

Central mechanisms—learned safety. The suggestion of an "active decay" process raises still another possibility, which is not a decay mechanism at all. Perhaps the delay gradient should be viewed not as a forgetting curve but as a learning curve. That is, in the absence of unfavorable gastrointestinal events, as time passes following consumption of a novel solution, the animal *learns that the solution is safe* (Kalat & Rozin, see Footnote 6).

As evidence for this, it has been demonstrated that an animal does learn something about a solution when consumption of that solution is followed by no negative consequences. We have already discussed the evidence indicating that rats learn less aversion to familiar solutions than to novel solutions (Revusky & Bedarf, 1967) even when the familiar solution was experienced once for only 20 minutes, three weeks before poisoning (Kalat & Rozin, see Footnote 6). This interpretation goes on to make a stronger claim: The learning of "safety" takes place within the period measured by the maximum delay of reinforcement. At the end of that period, the "trace" has not decayed; it has merely been associated with the *absence* of aversive consequences. At intermediate delays, the intermediate amount of learned aversion reflects the fact that the animal has learned an intermediate amount of safety. Although this interpretation is grossly different (and not inconsistent with) the interference or trace-decay interpretations, it is not easy to separate the alternatives experimentally. The results of Rozin and Ree (see Footnote 9), which suggest that indefinitely long CS–UCS intervals will support aversion learning if the rat is anesthetized during the interval, are consistent with learned safety, since it is reasonable that the rat would be unable to learn safety while anesthetized.

---

[9] P. Rozin and P. Ree. Long extension of effective CS–UCS interval by anesthesia between CS and US. In preparation.

WHAT ALL THIS MEANS TO THE RAT

We are now prepared to describe how a rat can handle some of the complex problems in food selection. The first thing to realize is that the situation is probably less complex than it might appear when the rat's natural behavior is considered. For example, the rat who gets sick in the garbage dump probably did not recently sample all the choice delicacies available (Rozin, 1969a). His choice behavior itself will help to unconfound the situation. He *may* have eaten a few different foods. He "knows" it was a *food* that made him sick (the belongingness principle) and can discount any familiar safe foods (the novelty principle). With the capability of forming associations over long delays, he is now likely to associate his illness with the last relevant (as defined above) thing or few things he ate over the last few hours. Although some of these foods may become more aversive than others because of their intrinsic properties (salience effect), the rat will acquire a significant aversion to each of them (small interference effect), with those closer in time to the aversive event picking up somewhat more aversion (temporal contiguity). Similar mechanisms would be employed to account for important aspects of the regulation of food intake.

What Kind of Learning Is This?

We have already described a number of differences between taste-aversion learning and traditional learning. We shall now consider some additional evidence in order to determine how fundamental this distinction is. Recent experiments have suggested two additional differences. First, it appears that rats can learn taste aversions when poisoned under anesthesia (Berger[10]; Roll & Smith, in press). Anesthesia was continued for a long period following UCS administration, so that it was highly likely that the learning occurred under anesthesia. Second, one of the more "complex" characteristics of classical conditioning, the Kamin blocking effect (Kamin, 1969), is either nonexistent or relatively weak in taste-aversion learning (Kalat & Rozin, 1971b; Revusky, in press).

One interpretation of why taste-aversion learning does not clearly show the Kamin-type effects is that the taste does not become a *signal* for poison in the sense that a tone or light becomes a signal for shock. In taste-aversion learning, the animal's perception of the taste itself or of its affective value may change (suggested by H. Gleitman, personal communication, 1971). The taste itself may become aversive or unacceptable, as if it were unpalatable (Rozin, 1967b). By contrast, stimuli associated with shock do not themselves become aversive; they evoke little avoidance outside the training situation. This difference seems related to the reported "*irrationality*" of learned taste aversions; humans commonly maintain an aversion to foods they ate prior to becoming nauseous, even when they are sure that some other agent was responsible for the nausea.

Seligman (1970) has suggested that taste-aversion learning may be a particularly primitive type of learning. The evidence mentioned above is quite consistent with this view. Furthermore, in a biological context, it makes sense that this ability should be primitive. The problems of food selection and the regulation of food intake are pervasive ones, and both are likely to involve long-delay learning. The essential problems are relatively invariant across species: any new food has some probability of being toxic, and for almost all species, the caloric density of foods will vary. Therefore, it makes sense that taste aversion and related types of learning should be more or less the same throughout most of the vertebrate subphylum.

---

[10] B. Berger. Learning in the anesthetized rat. In preparation.

ADAPTIVE SPECIALIZATIONS OF LEARNING: GENERALITY AND RELATION TO OTHER POSITIONS

The work on specific hungers and poisoning suggests that there are two aspects of adaptive specializations of learning. First, some mechanisms of learning may differ markedly (at least in terms of large quantitative differences in basic parameters) in the feeding system, as compared to other systems. Second, understanding of an adaptive specialization includes delimitation of the situations in which it applies and elaboration of its relationship and interaction with the animals' natural behavior (e.g., sampling, neophobia) in the relevant situation. The variation in interplay among naturalistic context, genetic programming, and learning is clearly illustrated in the contrast between water or sodium hunger, and other specific hungers.

Our emphasis has been on the first aspect of adaptation, while the ethologists have focused more on the context of learning and interplay of prestructured and acquired components:

> The student of innate behavior, accustomed to studying a number of different species and the entire behavior pattern, is repeatedly confronted with the fact that an animal may learn some things much more readily than others. . . . In other words, there seem to be more or less strictly localized "dispositions to learn." Different species are predisposed to learn different parts of the pattern. So far as we know, these differences between species have adaptive significance [Tinbergen, 1951, p. 145].

> . . . innumerable observations and experiments tend to show that modifiability occurs, if at all, only in those performed places where built-in learning mechanisms are phylogenetically programmed to perform just that function. How specifically these mechanisms are differentiated for one particular function is borne out by the fact that they are very often quite unable to modify any but one strictly determined system of behavior mechanisms [Lorenz, 1965, p. 47].

We are aware of only two well-studied systems showing adaptive specializations of learning. One is feeding, which we have already discussed, and the other involves imprinting, as a mechanism of species recognition.

Two of the prominent features of imprinting—the special sensitivity during a critical period early in life and the great resistance to extinction—can be seen as adaptations to limit the likelihood of errors in species recognition. The learning should take place soon after hatching, since the probability of exposure to a conspecific (i.e., the mother) is highest at this time, and the bird should be less sensitive to later experiences, since the frequency of contact with members of other species is likely to increase greatly after the nestling period.

Feeding and imprinting can be considered as two exceptions to an otherwise correct "general process" view of learning, or they can be considered as examples of a basic adaptational principle pervading much or all of learning. We prefer the latter alternative and believe that the absence of additional known instances of adaptive specializations may reflect learning psychologists' reluctance to study potentially learned behaviors which do not fit into the general process paradigms. (Significantly, the case for adaptive specializations in both imprinting and feeding came from outside the psychology of learning.)

At this point one can only speculate about what other systems will show special adaptations of learning. Bees may possess a wide variety of adaptive specializations (von Frisch, 1953, 1967). A particularly promising example is their navigational ability. First, there is evidence for "belongingness."

> Honey bees can learn to use irregular forms, like those of trees or rocks, as landmarks by which to steer a course to and from the hive; but they cannot,

> even by the most subtle conditioning technique, be taught to use the same forms as positive or negative signals indicating the presence or absence of food in a tray, as von Frisch (1914) has shown. As signals for food, bees can distinguish different forms only if they are geometrically regular, preferably radially symmetrical (Hertz, 1937) [Lorenz, 1965, p. 47].

Second, a limited amount of experience in observing a piece of the sun's arc enables the bee to "project" the rest of the arc. That is, bees raised without the opportunity to see the sun have great difficulty compensating for sun movement during navigation. But bees that have observed sun movement only during a limited period in the morning are fully capable of compensating for sun movement in the afternoon (von Frisch, 1967). In this situation, the environmental input produces a long-term change, and does not act "associatively," but rather provides a reference point. Other examples of such "calibrational learning" (Lorenz, 1965) might include adaptation to visual displacement produced by prisms and caloric regulation.[11]

The naturalistic literature is replete with other examples of surprising abilities of animals—such as digger wasps' memory of the location and state of their nests (Baerends, 1941), gobies' latent learning of the location of tidepools (Aronson, 1951), salmon's recognition of home-stream odor (Hasler, 1966), doves' individual mate recognition (Morris & Erickson, 1971), sparrows' acquisition and storage of song dialects (Marler, 1970), etc. These have not been thoroughly studied as possible specialized learning mechanisms.

The "heuristic" value of an adaptive evolutionary point of view can be suggested by considering the types of predictions that might be made about some basic learning and memory relationships. For example, an organism's memory of some aspect of the environment is useful only if that aspect is predictable or controllable; otherwise, rapid forgetting might be advantageous. It is probably of no use for bees, birds, or other organisms with complex navigational abilities to remember whether it was cloudy *yesterday* or which way the wind was blowing, though both may have been important at the time. Under these circumstances one might expect to find a specialization in short-term memory such that information could be stored for longer periods and in greater quantities than usual without entering long-term memory.

Another example concerns extinction. Extinction, from an adaptive point of view, allows an organism (*a*) to correct for mistakes (fortuitous conditioning) and (*b*) to constantly reshape itself to adapt to a variable environment. One might expect the rapidity of extinction to depend on the probability that either of these events would occur. In the case of imprinting, we have great resistance to extinction in a case where clearly the environment will not vary (i.e., the species will not change), and the proper imprinting object is almost certain to be present at the time of imprinting. The great resistance to extinction of avoidance conditioning (Solomon & Wynne, 1954) may be an adaptively selected feature of this learning: the costs of errors of omission here are high. One might expect rapid extinction of the learned location and stimulus properties of food sources, where these sources are subject to marked seasonal or other temporal variations. An example might be extinction of responses to particular types and locations of flowers in foraging bees. One might observe slow learning, but high resistance to extinction in a situation where the environment is stable but a critical difference is relatively hard to detect (e.g., prey density), and "averaging" must be done before a clear choice or preference is established. One might observe rapid learning and rapid extinction in a rapidly changing environment (see Shettleworth, 1971, for additional comments on this point).

---

[11] We can suggest a role for "calibrational learning" in the regulation of food intake. It has been known for some time that rats and other mammals respond to changes in caloric density of food by appropriate modulation of volume intake. In the rat, this compensation occurs largely as an increase in meal size, rather than in number of meals (Snowdon, 1969; Teitelbaum & Campbell, 1958). When a standard diet is diluted, the rat ends up eating larger meals: he has "recalibrated" his meal size on the basis of the metabolic aftereffects of his meals. It is possible to imagine a mechanism which compares some measure of the amount ingested with its delayed metabolic effects, and adjusts future intake downward or upward so that the metabolic consequences will approach some preferred or ideal value. The demonstration of a long-delay learning mechanism makes this type of explanation feasible and subject to investigation.

Recently, a number of authors have expressed positions related to ours (Garcia & Ervin, 1968; Garcia et al., 1971; Revusky, in press; Revusky & Garcia, 1970; Seligman, 1970). We differ from all of these authors in the sense that they see belongingness, in one form or another, as *the* unique phenomenon to be explained, whereas we see it as an example of the general adaptational principle; animals may not only learn some things more easily than others, but they may also learn some things in a different way than others. The contrast is clearest in the case of Revusky, who attempts to preserve traditional learning theory intact, with the introduction of a new "belongingness" assumption.

Seligman (1970) proposes a new dimension, "preparedness," based on the belongingness relationship, to incorporate recent findings into a more viable learning theory. Preparedness represents the tendency of certain inputs (and/or outputs) to be associated with one another, this tendency resulting from natural selection. Seligman proposes that highly prepared associations are established with a minimal input (e.g., number of pairings). In addition to very rapid learning, prepared associations would tend to show learning with long delays of reinforcement and perhaps high resistance to extinction. We are certainly in accord with the general flavor of Seligman's position, but we feel that in his desire to reorder the phenomena of learning, he has not fully appreciated the diverse natural forces that can shape behavior and learning mechanisms. In imprinting, for example, where Seligman's view of rapid learning (and high resistance to extinction) fits very well, the presence of the critical period cannot be accommodated without additional assumptions. More significantly, the long-delay learning found in feeding is probably not characteristic of other "prepared" associations, and in our view, it should not be, since in most cases close temporal contiguity is the best predictor. Furthermore, we see no reason to expect a consistent relationship between rapid learning and high resistance to extinction (see discussion of extinction above).

In short, we disagree with Seligman that diversity can be ordered along any single, operationally meaningful dimension, that is, preparedness. If preparedness meant adaptedness to situational demands, it would be an acceptable but not clearly meaningful dimension. Seligman has given it operational meaning, but narrowed the scope of the phenomena he can account for in the process. In a more recent statement of the preparedness position, Seligman and Hager (in press) have acknowledged some of these possible limitations.

Shettleworth (1971), in a paper written concurrently with this one, has presented a position very similar to ours. It focuses on belongingness, broadly conceived, but also describes learning in general in terms of adaptive specializations, and provides examples from the naturalistic literature. A recent paper by Lockard (1971) also explicitly discusses the diversity of learning mechanisms as an important feature ignored by most psychologists and implies that the search for common elements in learning across species and situations is almost hopeless.

We differ slightly from Shettleworth and markedly from Lockard with respect to our optimism about the possibility of finding order within the diversity of learning mechanisms. Given the constraints on adaptations produced by basic properties of the nervous system, the cost of evolving specializations, and the fact that most species face a common set of problems, we doubt that a separate learning mechanism would exist for every situation, or that there would be separate laws for each species. It may yet be possible to formulate laws of some degree of generality, taking ecological factors into account (see

Walls, 1942, for an example in the area of vision).

SOME SPECULATIONS

It follows from the point of view presented here that an organism may have an ability that manifests itself in only a small number of the total possible situations in which an experimenter might test for it. The ability in question might be inaccessible to or "unconnected" with the machinery for modulating and controlling most behavior. Using a computer analogy, we might suppose that a particular routine is designed (evolved) to handle a specific problem. At this point in time, it is physically connected only into the relevant inputs, outputs, or systems, and is inaccessible to the rest. Under these circumstances, it would be difficult to describe a species' learning capacity.

The interesting demonstrations by Bitterman (1968), Gonzalez and Bitterman (1969), and Mackintosh (1965) of differences in learning abilities in certain situations between a few species of fish and mammals do not in themselves indicate complete absence of these abilities in any species of fish (see Gleitman & Rozin, in press, for a general review of this issue with respect to learning and memory in fish). A species should be tested for any ability in those situations where its existence would have the greatest survival value. The failure of a few fish species to demonstrate abilities such as progressive improvement in habit reversal in a few laboratory situations does not meet this demand.

However, the data gathered by Bitterman, by Gonzalez and Bitterman, and by Mackintosh do suggest some interesting phylogenetic generalizations about learning capacities. Since the laboratory apparatus used to study rats and other mammals is often not ideally suited to *their* natural behavior, just as the fish apparatus is not, it *is* quite interesting that mammals reliably show the greater plasticity in these "unnatural" situations. One possible explanation for this, offered by Bitterman (1964), is that the rat possesses certain higher learning abilities that "fish" do not possess. Another not mutually exclusive possibility is that the rat and probably most other mammals are "generalists," compared to the majority of other vertebrates. That is, abilities initially evolved to handle a specific situation and limited in application to the appropriate system may turn out to be useful in others and may through evolution be "connected" into new systems. To what extent do the higher abilities of the mammals represent an increase in accessibility of specialized capacities, an "emancipation," to borrow a word from the ethologists, of a capacity from its original tight motivational system?

The proposed increase in accessibility of capacities in phylogeny may have a parallel in ontogeny. Within the Piagetian framework, it is apparent that particular cognitive structures may have only limited applicability at any point in development. Piaget applies the name "decalage" to this feature of development. (Flavell, 1963, pp. 21–24). For example, in the Piagetian scheme, the same cognitive structure is necessary for the achievement of mass and weight conservation, yet the latter occurs about two years after the former. Cognitive development may consist, in part, of the extension of existent capacities to new situations, in parallel to our scheme for phylogeny.

We have argued for the existence of adaptive specializations in learning and memory. Since, by their nature, such specializations are limited to a relatively narrow range of situations, we have pointed out that they must be inaccessible to most functioning systems. We believe that this general formulation has a wide application, and we are presently applying it to an understanding of the difficulties in initial acquisition of reading (Rozin & Kalat, in press; Rozin, Poritsky, & Sotsky, 1971).

It is the basic thesis of this paper that in a biological-evolutionary framework, specifically adapted abilities are extremely important and should not suffer from neglect as a consequence of the search for great generalities. To understand a set of phenomena, within humans or across the animal kingdom, is to be able to describe and explain diversity, as well as to extract common elements.

REFERENCES

Appledorf, H., & Tannenbaum, S. R. Specific hunger for thiamine in the rat: Selection of low concentrations of thiamine in solution. *Journal of Nutrition*, 1967, **92**, 267–273.

Aronson, L. Orientation and jumping in the gobiid fish *Bathygobius soporator*. *American Museum Novitates*, 1951, **1486**, 1–22.

Baerends, G. P. Fortpflanzungsverhalten und orientierung der grabwespe *Ammophila campestris* Jur. *Tijdschr Entomol*, 1941, **84**, 68–275.

Barnett, S. A. *The rat. A study in behaviour*. London: Methuen, 1963.

Bitterman, M. E. The evolution of intelligence. *Scientific American*, 1964, **212**, 92–100.

Bitterman, M. E. Comparative studies of learning in the fish. In D. J. Ingle (Ed.), *The central nervous system and fish behavior*. Chicago: University of Chicago Press, 1968.

Bradley, R. M., & Mistretta, C. M. Intravascular taste in rats as demonstrated by conditioned aversion to sodium saccharin. *Journal of Comparative and Physiological Psychology*, 1971, **75**, 186–189.

Brower, L. P. Ecological chemistry. *Scientific American*, 1969, **220**, 22–29.

Bykov, K. M. *The cerebral cortex and the internal organs*. (Trans. & ed. by W. H. Gantt) New York: Chemical Publishing, 1957.

Campbell, C. Development of specific preferences in thiamine-deficient rats: Evidence against mediation by aftertastes. Unpublished master's thesis, Library of the University of Illinois at Chicago Circle, 1969.

Capranica, R. R. Vocal response of the bullfrog to natural and synthetic mating calls. *Journal of the Acoustical Society of America*, 1966, **40**, 1131–1139.

Capretta, P. J. An experimental modification of food preference in chickens. *Journal of Comparative and Physiological Psychology*, 1961, **54**, 238–242.

Flavell, J. H. *The developmental psychology of Jean Piaget*. Princeton: Van Nostrand, 1963.

Galef, B. G. Aggression and timidity: Responses to novelty in feral Norway rats. *Journal of Comparative and Physiological Psychology*, 1970, **70**, 370–381.

Galef, B. G. Social effects in the weaning of domestic rat pups. *Journal of Comparative and Physiological Psychology*, 1971, **75**, 358–362.

Galef, B. G., Jr., & Clark, M. M. Social factors in the poison avoidance and feeding behavior of wild and domestic rat pups. *Journal of Comparative and Physiological Psychology*, 1971, **75**, 341–357. (a)

Galef, B. G., Jr., & Henderson, P. W. Mother's milk: A determinant of the feeding preferences of weaning rat pups. *Journal of Comparative and Physiological Psychology*, in press.

Galef, B. G., Jr., & Clark, M. M. Mother's milk and adult presence: Two factors determining initial dietary selection by weanling rats. *Journal of Comparative and Physiological Psychology*, in press.

Garcia, J., & Ervin, F. R. Gustatory-visceral and telereceptor-cutaneous conditioning—Adaptation in internal and external milieus. *Communications in Behavioral Biology*, 1968, **1** (Pt. A), 389–415.

Garcia, J., Ervin, F. R., & Koelling, R. A. Learning with prolonged delay of reinforcement. *Psychonomic Science*, 1966, **5**, 121–122.

Garcia, J., Ervin, F. R., Yorke, C. H., & Koelling, R. A. Conditioning with delayed vitamin injection. *Science*, 1967, **155**, 716–718.

Garcia, J., Green, K. F., & McGowan, B. K. X-ray as an olfactory stimulus. In C. Pfaffman (Ed.), *Taste and olfaction*. New York: Rockefeller University Press, 1969.

Garcia, J., Kimeldorf, D. J., & Hunt, E. L. The use of ionizing radiation as a motivating stimulus. *Psychological Review*, 1961, **68**, 383–385.

Garcia, J., & Koelling, R. A. Relation of cue to consequence in avoidance learning. *Psychonomic Science*, 1966, **4**, 123–124.

Garcia, J., & Koelling, R. A. A comparison of aversion induced by x-rays, toxins, and drugs in the rat. *Radiation Research*, 1967, Suppl. 7, 439–450.

Garcia, J., McGowan, B. K., & Green, K. F. Biological constraints on conditioning. In A. Black & W. F. Prokasy (Eds.), *Classical conditioning*. New York: Appleton-Century-Crofts, in press.

Gibson, J. J. *The senses considered as perceptual systems*. Boston: Houghton Miffln, 1966.

Gleitman, H., & Rozin, P. Learning and memory. In W. F. Hoar & D. J. R. Randall (Eds.), *Fish physiology*. New York: Academic Press, in press.

Gonzalez, R. C., & Bitterman, M. E. The spaced-trials PRE as a function of contrast. *Journal of Comparative and Physiological Psychology*, 1969, **67**, 94–103.

Harper, A. E. Effects of dietary protein content and amino acid pattern on food intake and preference. In C. F. Code & W. Heidel (Eds.), *Handbood of physiology*. Vol. 1, Section 6. Washington, D. C.: American Physiological Society, 1967.

Harris, L. J., Clay, J., Hargreaves, F., & Ward, A. Appetite and choice of diet. The ability of the vitamin B deficient rat to discriminate between diets containing and lacking the vitamin. *Proceedings of the Royal Society, London (Series B)*, 1933, **113**, 161–190.

Hasler, A. D. *Underwater guideposts. Homing of salmon*. Madison: University of Wisconsin Press, 1966.

Hertz, M. Beitrag zum Farbensinn und Formensinn der Biene. *Zeitschrift für vergleichende Physiologie*, 1937, **24**, 413–421.

Hunt, G. L., & Smith, W. J. Pecking and initial drinking responses in young domestic fowl. *Journal of Comparative and Physiological Psychology*, 1967, **64**, 230–236.

Kalat, J. W., & Rozin, P. "Salience": A factor which can override temporal contiguity in taste-aversion learning. *Journal of Comparative and Physiological Psychology*, 1970, **71**, 192–197.

Kalat, J., & Rozin, P. Role of interference in taste-aversion learning. *Journal of Comparative and Physiological Psychology*, 1971, **77**, 53–58. (a)

Kalat, J., & Rozin, P. You can lead a rat to poison—but you can't make him think. In M. E. P. Seligman & J. Hager (Eds.), *Biological boundaries of learning*. New York: Appleton-Century-Crofts, 1971, in press. (b)

Kamin, L. J. Predictability, surprise, attention, and conditioning. In B. Campbell & R. Church (Eds), *Punishment and aversive behavior*. New York: Appleton-Century-Crofts, 1969.

Kimble, G. A. *Hilgard and Marquis' conditioning and learning*. New York: Appleton-Century-Crofts, 1961.

Kral, P. Interpolation of electroconvulsive shock during CS–US interval as an impediment to the conditioning of taste aversion. *Psychonomic Science*, 1970, **19**, 36–38.

Krieckhaus, E. E., & Wolf, G. Acquisition of sodium by rats: Interaction of innate mechanisms and latent learning. *Journal of Comparative and Physiological Psychology*, 1968, **65**, 197–201.

LeMagnen, J. Peripheral and systemic actions of food in the caloric regulation of intake. *Annals of the New York Academy of Science*, 1969, **157**, 1127–1157.

Lettvin, J. Y., Maturana, H. R., McCulloch, W. S., & Pitts, W. H. What the frog's eye tells the frog's brain. *Proceedings of the Institute of Radio Engineers*, 1959, **47**, 1940–1951.

Lockard, R. B. Reflections on the fall of comparative psychology: Is there a message for us all? *American Psychologist*, 1971, **26**, 168–179.

Lorenz, K. *Evolution and modification of behavior*. Chicago: University of Chicago Press, 1965.

Mackintosh, N. J. Selective attention in animal discrimination learning. *Psychological Bulletin*, 1965, **64**, 124–150.

Maier, S. F., Zahorik, D. M., & Albin, R. W. Relative novelty of solid and liquid diet during thiamine deficiency determines development of thiamine-specific hunger. *Journal of Comparative and Physiological Psychology*, 1971, **74**, 254–262.

Marler, P. A comparative approach to vocal learning: Song development in white-crowned sparrows. *Journal of Comparative and Physiological Psychology*, 1970, **71** (2, Pt. 2).

McLaurin, W. A. Postirradiation sacharin avoidance in rats as a function of the interval between ingestion and exposure. *Journal of Comparative and Physiological Psychology*, 1964, **57**, 316–317.

Morgan, C. L. *An introduction to comparative psychology*. London: Walter Scott, 1894.

Morris, R. L., & Erickson, C. J. Pair-bond maintenance in the ring dove (*Streptopelia risoria*). *Journal of Animal Behavior*, in press.

Nachman, M. Taste preferences for sodium salts by adrenalectomized rats. *Journal of Comparative and Physiological Psychology*, 1962, **55**, 1124–1129.

Nachman, M. Learned taste and temperature aversions due to lithium chloride sickness after temporal delays. *Journal of Comparative and Physiological Psychology*, 1970, **73**, 22–30. (a)

Nachman, M. Limited effects of electroconvulsive shock on memory of taste stimulation. *Journal of Comparative and Physiological Psychology*, 1970, **73**, 31–37. (b)

Pavlov, I. *Conditioned reflexes*. Oxford: Oxford University Press, 1927.

Rachman, S., & Teasdale, J. *Aversion therapy and behavior disorders: An analysis*. Coral Gables: University of Miami Press, 1969.

Revusky, S. H. Hunger level during food consumption: Effects on subsequent preference. *Psychonomic Science*, 1967, **7**, 109–110.

Revusky, S. H. Aversion to sucrose produced by contigent x-irradiation: Temporal and dosage parameters. *Journal of Comparative and Physiological Psychology*, 1968, **65**, 17–22.

Revusky, S. H. The role of interference in association over a delay. In W. Honig & H. James (Eds.), *Animal memory*. New York: Academic Press, in press.

Revusky, S. H., & Bedarf, E. W. Association of illness with prior ingestion of novel foods. *Science*, 1967, **155**, 219–220.

Revusky, S. H., & Garcia, J. Learned associations over long delays. In C. H. Bower & J. T. Spence (Eds.), *The psychology of learning and motivation: Advances in research and theory, IV*. New York: Academic Press, 1970.

Richter, C. P. Increased salt appetite in adrenalectomized rats. *American Journal of Physiology*, 1936, **115**, 155–161.

Richter, C. P. Total self-regulatory functions in animals and human beings. *Harvey Lectures Series*, 1943, **38**, 63–103.

Richter, C. P. Experimentally produced behavior reactions to food poisoning in wild and domesticated rats. *Annals of the New York Academy of Science*, 1953, **56**, 225–239.

Richter, C. P. Salt appetite of mammals: Its dependence on instinct and metabolism. In, *L'Instinct dans le comportement des animaux et de l'homme*. Paris: Masson, 1956.

Richter, C. P., Holt, L. E., Jr., & Barelare, B., Jr. Vitamin B_1 craving in rats. *Science*, 1937, **86**, 354–355.

Rodgers, W. L. Specificity of specific hungers. *Journal of Comparative and Physiological Psychology*, 1967, **64**, 49–58. (a)

Rodgers, W. L. Thiamine specific hunger. Unpublished doctoral dissertation, University of Pennsylvania, 1967. (b)

Rodgers, W., & Rozin, P. Novel food preferences in thiamine-deficient rats. *Journal of Comparative and Physiological Psychology*, 1966, **61**, 1–4.

ROEDER, K. *Nerve cells and insect behavior.* Cambridge: Harvard University Press, 1963.

ROGERS, Q. R., & HARPER, A. E. Selection of a solution containing histidine by rats fed a histidine-imbalanced diet. *Journal of Comparative and Physiological Psychology*, 1970, **72**, 66–71.

ROLL, D. L., & SMITH, J. C. Conditioned taste aversion in anesthetized rats. In M. E. P. Seligman & J. Hager (Eds.), *The biological boundaries of learning.* New York: Appleton-Century-Crofts, in press.

ROTHSCHILD, M. Mimicry. The deceptive way of life. *Natural History*, 1967, **76**, 44–51.

ROZIN, P. Specific hunger for thiamine: Recovery from deficiency and thiamine preference. *Journal of Comparative and Physiological Psychology*, 1965, **59**, 98–101.

ROZIN, P. Thiamine specific hunger. In, *Handbook of physiology.* Vol. 1, Section 6. Washington, D. C.: American Physiological Society, 1967. (a)

ROZIN, P. Specific aversions as a component of specific hungers. *Journal of Comparative and Physiological Psychology*, 1967, **64**, 237–242. (b)

ROZIN, P. Specific aversions and neophobia as a consequence of vitamin deficiency and/or poisoning in half-wild and domestic rats. *Journal of Comparative and Physiological Psychology*, 1968, **66**, 82–88.

ROZIN, P. Adaptive food sampling patterns in vitamin deficient rats. *Journal of Comparative and Physiological Psychology*, 1969, **69**, 126–132. (a)

ROZIN, P. Central or peripheral mediation of learning with long CS-US intervals in the feeding system. *Journal of Comparative and Physiological Psychology*, 1969, **67**, 421–429. (b)

ROZIN, P., & KALAT, J. W. Learning as a situation-specific adaptation. In M. E. P. Seligman & J. Hager (Eds.), *The biological boundaries of learning.* New York: Appleton-Century-Crofts, in press.

ROZIN, P., PORITSKY, S., & SOTSKY, R. American children with reading problems can easily learn to read English represented by Chinese characters. *Science*, 1971, **171**, 1264–1267.

ROZIN, P., & RODGERS, W. Novel diet preferences in vitamin deficient rats and rats recovered from vitamin deficiencies. *Journal of Comparative and Physiological Psychology*, 1967, **63**, 421–428.

ROZIN, P., WELLS, C., & MAYER, J. Thiamine specific hunger: Vitamin in water versus vitamins in food. *Journal of Comparative and Physiological Psychology*, 1964, **57**, 78–84.

RŹOSKA, J. Bait shyness, a study in rat behaviour. *British Journal of Animal Behaviour*, 1953, **1**, 128–135.

SCOTT, E. M., & QUINT, E. Self selection of diet. III. Appetites for B vitamins. *Journal of Nutrition*, 1946, **32**, 285–291.

SCOTT, E. M., & VERNEY, E. L. Self selection of diet. VI. The nature of appetites for B vitamins. *Journal of Nutrition*, 1947, **34**, 471–480.

SCOTT, E. M., VERNEY, E. L., & MORISSEY, P. Self selection diet. XI. Appetites for calcium, magnesium and potassium. *Journal of Nutrition*, 1950, **41**, 187–202.

SELIGMAN, M. E. P. On generality of the laws of learning. *Psychological Review*, 1970, **77**, 406–418.

SELIGMAN, M. E. P., & HAGER, J. *The biological boundaries of learning*, New York: Appleton-Century-Crofts, in press.

SHETTLEWORTH, S. Constraints on learning. In D. S. Lehrman, R. A. Hinde, & E. Shaw (Eds.), *Advances in the study of behavior.* Vol. 4. New York: Academic Press, 1971.

SMITH, J. C., & ROLL, D. L. Trace conditioning with X-rays as an aversive stimulus. *Psychonomic Science*, 1967, **9**, 11–12.

SMITH, J. C., & SCHAEFFER, R. W. Development of water and saccharin preferences after simultaneous exposures to saccharin solution and gamma rays. *Journal of Comparative and Physiological Psychology*, 1967, **63**, 434–438.

SMITH, M. P., & CAPRETTA, P. J. Effects of drive level and experience on the reward value of saccharine solutions. *Journal of Comparative and Physiological Psychology*, 1956, **49**, 553–557.

SMITH, R. G. The role of alimentary chemoreceptors in the development of taste aversions. *Communications in Behavioral Biology*, 1970, **5**, 199–204.

SNOWDON, C. T. Motivation, regulation, and the control of meal parameters with oral and intragastric feeding. *Journal of Comparative and Physiological Psychology*, 1969, **69**, 91–100.

SOLOMON, R. L., & WYNNE, L. C. Traumatic avoidance learning: The principles of anxiety conservation and partial irreversibility. *Psychological Review*, 1954, **61**, 353–385.

STRICKER, E. M., & STERRITT, G. M. Osmoregulation in the newly hatched domestic chick. *Physiology and Behavior*, 1967, **2**, 117–119.

STRICKER, E. M., & WILSON N. E. Salt-seeking behavior in rats following acute sodium deficiency. *Journal of Comparative and Physiological Psychology*, 1970, **72**, 416–420.

TEITELBAUM, P., & CAMPBELL, B. A. Ingestion patterns in hyperphagic and normal rats. *Journal of Comparative and Physiological Psychology*, 1958, **51**, 135–141.

TINBERGEN, N. *The study of instinct.* Oxford: Clarendon Press, 1951.

VON FRISCH, K. Der Farbensinn und Formensinn der Biene. *Zoologisches Jahrbuch*, 1914, **35**, 1–188.

VON FRISCH, K. *The dancing bees.* New York: Harcourt, Brace, 1953.

VON FRISCH, K. *The dance language and orientation of bees.* Cambridge, Mass.: Belknap Press of Harvard University Press, 1967.

WALLS, G. L. *The vertebrate eye and its adaptive radiation.* Bloomfield Hills, Mich.: Cranbrook Institute of Science, 1942.

WICKLER, W. *Mimicry in plants and animals.* New York: McGraw-Hill, 1968.

WILCOXON, H. C., DRAGOIN, W. B., & KRAL, P. A. Illness-induced aversions in rat and quail: Relative salience of visual and gustatory cues. *Science,* 1971, **171**, 826–828.

WITTLIN, W. A., & BROOKSHIRE, K. H. Apomorphine-induced conditioned aversion to a novel food. *Psychonomic Science*, 1968, **12**, 217–218.

ZAHORIK, D. M., & MAIER, S. F. Appetitive conditioning with recovery from thiamine deficiency as the unconditioned stimulus. *Psychonomic Science*, 1969, **17**, 309–310.

(Received September 15, 1970; revision received July 12, 1971)

Psychological Review
1971, Vol. 78, No. 6, 487–504

ADJECTIVES, COMPARATIVES, AND SYLLOGISMS[1]

JANELLEN HUTTENLOCHER[2] AND E. TORY HIGGINS

Teachers College, Columbia University

This paper outlines a more complete explanation of how people solve ordering syllogisms than has previously been given. Earlier, one of the authors proposed that people combine information from the two syllogistic premises using a strategy which involves spatial imagery. A more complete explanation of the total process of solving these syllogisms would include an account of the comprehension of the adjectives and comparative sentences which must be understood to solve the problems, as well as an account of question answering. The present paper is concerned with these issues and, in addition, with an evaluation of the relative explanatory value of the spatial imagery proposal as against more abstract formulations.

The present paper is concerned with the mental processes involved in solving ordering syllogisms or "three-term series" problems. These problems typically include two comparative sentences which specify the order of three items on some dimension, plus a question about order; for example, "Tom is taller than Harry. Joe is shorter than Harry. Who is tallest?" Since the information about item order is given in two separate sentences, each of which describes the relation between two of the items, people must put together information from these separate sentences to determine the order of all three items. In order to put that information together, they must be able to understand the comparative sentences, and to do this, in turn, they must understand the adjectives within these sentences. Finally, in order to solve the problem, they must also understand the question.

In explaining the mental processes of the problem solver, one may deal only with the issue of how information from the two premises is combined, as Huttenlocher (1968) did. This process is traditionally what is meant by reasoning, as separable from merely understanding the individual premises. A more complete explanation of the mental processes people use in getting from the statement of the problem to the answer would, in addition, include an account of the comprehension of adjectives and comparatives as well as an account of question answering. In an earlier paper, Clark (1969) attempted to do the latter, while basically ignoring the question of how information from the two premises is combined. In this paper, we consider all three aspects of the problem-solving process.

ADJECTIVES

Adjectives fall into various semantic classes (e.g., height, color). The members of such classes (e.g., for height: tall, short; for color: red, yellow; etc.) have many elements of meaning in common. Such semantic classes can be grouped into different types which vary with respect to how the member adjectives are related to one another. Our concern in this paper is with the type of semantic class where the member adjectives fall into single continuous dimensions allowing for the ordering of various items within them. Examples are tall–short, beautiful–ugly, hot–cold, etc. Such dimensions account for a large number of adjectives and can be contrasted with somewhat different adjective classes—dichotomous classes (e.g., dead–alive, male–female, etc.) and classes where the members cannot fit on a single dimension (e.g., shape, with the members circular, square, etc.; color, with the members red, yellow, etc.).

---

[1] The preparation of this paper was supported in part by Career Development Award 5-K-4-HD-21,979, and in part by Research Grant HD 03215, both from the National Institutes of Health to Janellen Huttenlocher.

[2] Request for reprints should be sent to Janellen Huttenlocher, Teachers College, Columbia University, New York, New York 10027.

Different Kinds of Continuous Dimensions

There are many kinds of continuous dimensions, involving all those ways people rank objects, persons, and events with respect to particular properties, qualities, and values. One kind of dimension which is easily distinguishable from others is the kind which describes quantitative properties of material objects and of space, time, etc.; for example, length (long–short), width (wide–narrow), speed (fast–slow). The domain described by the adjectives in such dimensions usually involves a metric (inches, pounds, miles per hour, etc.) as well as a true zero point (no length, weight, or speed). Thus the dimension describes a ratio scale, and one can make statements like "John is five feet tall," "Building A is twice as high as Building B."

Outside this set of metric dimensions, many of which are spatial, the largest number of other continuous dimensions lack both a metric and a zero point, thus forming only ordinal scales. Therefore, items can only be rank ordered within them. Included in this set are various dimensions of value and quality; for example, beauty (beautiful–ugly), kindness (kind–unkind), happiness (happy–sad), and evaluation (good–bad). There are, in addition, some dimensions which lack a true zero point but nevertheless allow one to consider relative differences between different pairs of points, thus forming interval scales; for example, temperature (hot–cold) and intelligence (intelligent–unintelligent).

Metaphoric usage of adjectives is common, and one cannot necessarily conclude that an adjectival expression involving a spatial dimension is necessarily being used in its mundane measurement sense; for example, "He felt twice as tall after winning the award," "He has a shallow knowledge of foreign affairs." It also must be pointed out that the above division of adjectives into semantic sets of certain characteristics does not cover every case. At least one spatial dimension, that of overall size as encoded in the terms "big" and "small," has no metric although it surely has a zero point. A dimension which would seem to have an obvious zero point and an obvious metric, that of wealth ("rich" or "poor"), is not often used in that way ("He is $250,000 rich" is not an acceptable sentence) but rather is used chiefly as an ordinal scale which seems to reach out unbounded in both directions.

Marked and Unmarked Adjectives

The polar adjectives for the two ends of most continuous dimensions are asymmetrical in various ways; one of these polar adjectives is typically designated as "marked" and the other as "unmarked." In his analysis of spatial adjectives (a category which is largely synonymous with our category of ratio adjectives), Bierwisch (1967) has proposed that the meanings of such polar pairs might be broken up into various elements, most of which the two adjectives share, but on one of which they differ, namely, which pole each designates. He suggested assigning to one adjective a "+" and to the other a "−." The former is typically called the unmarked and the latter the marked adjective.

For the spatial adjectives he has dealt with, there are two bases for deciding which adjective should be assigned a "+": First, the unmarked adjective indicates the presence of a property which can extend indefinitely in an upward direction. The other adjective, which is assigned the "−" pole, indicates the absence of that property, the extreme lower bound being zero. To be tall, wide, etc., is to possess something, whereas to be short, narrow, etc., is to lack rather than to have something. Second, as Bierwisch (1967) and Clark (1969) have pointed out, while both the marked and the unmarked adjective may be used contrastively to attribute to an item or class a particular position along the dimension, (e.g., "A is tall," "A is wide," "A is narrow," "A is short") the unmarked adjective has, in addition, a nominal usage, for which there is no parallel with the marked adjective. That is, the unmarked adjective can also be used to refer to the entire dimension (e.g., "How tall is the girl?" "How wide is the road?"). To use the marked adjective in this way is either anomalous or seems to imply a prior judgment about the item one is referring to, as in "How short is the girl?"

These two sources of asymmetry between the polar adjectives from ratio scales are obviously related. That is, a dimension designates a property which can vary in extent. Since the unmarked adjective designates the presence of the property, whereas the marked adjective designates its absence, the former would seem the more sensible way to name the dimension than the latter. That is, there are units of length, width, etc., but not of shortness, narrowness, etc.

The dimension name is not always a form of the unmarked adjective; for example, height, speed, etc. Nevertheless, for these dimensions as well, the unmarked adjective can be used in the same neutral sense as when these words do coincide, whereas the marked adjectives cannot; for example, one can neutrally say "How tall is the girl?" or "How fast can he run?" but not "How short is the girl?" or "How slow can he run?"

For adjectives from ordinal scales, the asymmetry between marked and unmarked adjectives does not derive from the dimension having a zero point and extending indefinitely in an upward direction. Rather, most such scales seem to extend indefinitely in *both* directions, as in good–bad and happy–sad. Therefore, the marked adjectives from such scales do not literally indicate complete absence of a property defined by the unmarked adjective. That is, whereas narrow means the absence of the property of width, sad does not so much mean the absence of happiness as it means the presence of a property of sadness. Nevertheless, as is shown below, people do treat the marked adjectives from ordinal scales as if they designated the absence of the property exhibited by the unmarked adjective.

The other basis for the asymmetry between the marked and unmarked adjectives for ratio scales was that the unmarked adjective could be used to refer to the entire dimension. This is true of the unmarked adjective for ordinal scales as well, as in "How happy is he at his new job?" or "How good is that restaurant?" although it could be argued that the unmarked adjective from ordinal scales does not cover the whole dimension in the same sense as the unmarked adjective from ratio scales. That is, as was said above, sadness is in some sense a property in itself rather than just the absence of happiness. As with ratio scales, the marked adjective from ordinal scales cannot be used nominally in a neutral sense; for example, "How sad is he at his new job?"

Given that ordinal scales, unlike ratio scales, extend in both directions from what might be called a zero point, one could in some sense regard them as separate dimensions; for example, happiness and sadness. It is not clear, therefore, why one of the adjectives is chosen over the other to be used neutrally. One common factor uniting those adjectives which can be used neutrally is that they designate the affectively or evaluatively positive poles of these ordinal scales; for example, kind, happy, good, etc.

Not only is the affectively positive adjective from ordinal scales used neutrally, and taken to indicate the presence of a property, but, conversely, the unmarked adjective from ratio scales is quite commonly used to indicate positive affect, as in such metaphoric phrases as "broad interests." Similarly, the marked adjective from ratio scales can be used metaphorically to indicate negative affect, as in "low down deal," "short end of a bargain," etc.

For both ordinal and ratio scales, the dimension designated by the polar pair, x and y, includes a region around the midpoint between the two poles where an item is neither "x" nor "y" but neutral. "Not-x" includes not only "y" but this neutral region. Even those marked adjectives which are formed by adding a negative prefix to their unmarked counterparts do not simply mean "not-x"; it is possible to be "not wise" without being "unwise," or to be "not kind" without being "unkind." On either side of this neutral region, an item is "x" or "y." In this sense, "x" and "y" are opposites.

The meanings of the marked and unmarked polar opposites differ in complexity, in that the meaning of the marked adjective seems to be built on the meaning of the unmarked adjective plus a logical operation of opposition. That is, "John is short" seems to mean "John is the opposite of tall." Furthermore, the logical operation by which the meaning of the marked adjective is built upon the meaning of its unmarked counterpart is in some sense negative. That is, if "x" is the unmarked and "y" the marked adjective, "y" seems to mean "no x." "No x" here includes only the marked end of the dimension, in contrast to the negation of "x," "not-x," which includes all portions of the dimension except the "x" portion. For ratio scales, the marked adjective literally designates the absence of a property, and the neutral region designates some average amount of the property. For ordinal scales, as was noted above, one could argue that such an interpretation of the scale is only metaphorical. Nevertheless, the evidence for the logical negativity of the marked adjectives involves both ratio and ordinal scales and comes from several sources: from the forms of definitions given for both types of adjectives, from internal linguistic evidence, and from differences in comprehension difficulty.

Mann (1968) reports an experiment where subjects (*S*s) were asked to define certain "favorable" and "unfavorable" adjectives. Generally, the "favorable" and "unfavorable" adjectives come from ordinal scales and could be classified as "unmarked" and "marked" adjectives, respectively. The *S*s were more likely to define "unfavorable" than "favorable" adjectives by denial. The vast majority of the denial definitions for "unfavorable" adjectives involved denial of the presence of their "favorable" opposites.

We did an analogous study specifically involving marked and unmarked adjectives from both ratio and ordinal scales. We extracted definitions from dictionaries in order to compare the extent to which marked adjectives were defined as negations of the unmarked form and vice-versa. Twenty pairs of common polar adjectives were chosen (see Table 1). Of these 40 adjectives, 31 were in the highest frequency category (AA) of the Thorndike and Lorge (1944) word frequency count. We randomly assigned these adjective pairs to four groups of five pairs each, and then assigned each group to one of four major dictionaries—Webster, Random-House, Oxford, and Funk and Wagnall. To make our analysis, we used the first dictionary definition which was specifically applicable to the dimension we were considering. If this first definition was stated in terms of the negative of the other member of the pair, we gave it one point. Otherwise, it was given a zero. We set a very stringent criterion for negativity, excluding even those adjectives defined relative to the norm or standard of the dimension using words such as "less than," "having little." For instance, "bad" was defined as "not good" and received one point. "Small" was defined as "comparatively less than another or than a standard" and received a zero. Not one of the unmarked adjectives was defined as the negation of the marked member of the pair; 12 of the 20 marked adjectives from both ordinal and ratio scales were defined as the negation of the

TABLE 1

TWENTY PAIRS OF POLAR ADJECTIVES FOR WHICH DICTIONARY DEFINITIONS WERE OBTAINED

Unmarked	Marked
big	small
hot	cold
high	low
noisy	quiet
deep	shallow
good	bad
rich	poor
tall	short
wide	narrow
dirty	clean
kind	cruel
happy	sad
old	young
heavy	light
fast	slow
strong	weak
beautiful	ugly
hard	soft
fat	thin
smart	stupid

unmarked member of the pair. Thus marked adjectives are often defined in terms of "lacking" or "wanting" in the property designated by the unmarked member of the bi-polar pair. Additional evidence that the marked adjective is thought in terms of the unmarked plus some additional operation, but not the reverse, is found in a memory experiment by Clark and Card (1969) where *S*s more often recalled a marked adjective as unmarked than vice-versa.

Another type of evidence about the negativity of marked adjectives comes from linguistic data showing that marked adjectives function in some ways like negatively prefixed adjectives. Klima (1964) discusses sentence frames which are acceptable only when a negatively prefixed word is inserted in the blank. He notes that certain "adversatives" or "inherent negatives" function like negatively prefixed words. At least some marked adjectives seem to be adversatives in this sense, since they fit meaningfully into sentences with "any," which require a negative element, whereas their unmarked counterparts do not. For example, (*a*) He would be stupid (*smart) to take any more. (*b*) It was cruel (*kind) to do any more. (*c*) He was sad (*happy) to see any more conflict.

Additional linguistic evidence also suggests that marked adjectives from both ratio and ordinal scales may incorporate an implicit negative. Zimmer (1964) points out that negative prefixes are rarely attached to the marked adjective, whereas they are commonly attached to the unmarked form; for example, unhappy, unattractive, but not *unsad, *unugly, etc. If the marked adjective contains an implicit negative, then adding a negative prefix would result in a double negative, which one would expect people to avoid, as they generally avoid expressions with double negatives. For instance, as "sad" is defined as "unhappy," "unsad" would mean *"ununhappy."

It might seem reasonable at first glance to give an alternative argument—that it is the negative affect of the marked adjective rather than its logical negativity which makes it unacceptable to add the negative prefix. This seems unlikely, however, since there are examples of adjectives with negative affect which *do* take a negative prefix: unspoiled, unbroken, uninjured, etc.

Further evidence of the logical negativity of marked adjectives comes from a psychological study by Sherman (1969). He showed that negative sentences with marked adjectives are more difficult to comprehend than identical sentences with their unmarked counterparts. The addition of a marked adjective increased the difficulty of sentence comprehension in much the same way as the addition of another explicit negative did. As is shown below, it is also more difficult to comprehend comparative expressions involving marked adjectives than those involving their unmarked counterparts.

An alternative explanation for the difference in the difficulty of marked and unmarked adjectives is that the unmarked adjective is often more frequent than its marked counterpart. However, there is some evidence that frequency is not the determining factor. In a few cases, the marked adjective is more frequent than the unmarked adjective, according to the Thorndike-Lorge frequency count ("short" and "poor" are more frequent than "tall" and "rich," respectively) and, for several other pairs, the frequencies of marked and unmarked pairs are equal. The results for such pairs parallel the results for other pairs (Huttenlocher, 1968; also Huttenlocher[3]). In fact, one reason marked adjectives are generally less frequent than unmarked adjectives may be because they are negative and thus more complicated, rather than vice-versa.

Other Adjectives from Continuous Dimensions

For certain dimensions, there are in addition to the two adjectives which can be designated as marked and unmarked, other adjectives which are not simply synonyms of these two polar adjectives. For example, with respect to size there are, in addition to the polar pair "big–small," the terms "tiny," "huge," miniscule," and "gigantic." With respect to human girth there are, in addition to the polar pair "fat–thin," the terms "obese," "skinny." For temperature there are two commonly used polar pairs; namely, "warm–cool," and "hot–cold," and, in addition, terms like "tepid" and "lukewarm." Many of these additional adjectives, like "huge" and "tiny," refer to portions of the dimension more extreme than those ordinarily used to designate the poles. We will have reason to discuss pairs of extreme adjectives of opposite meaning below when we consider the synonymy of different types of expressions involving pairs of polar adjectives.

---

[3] J. Huttenlocher. On answering questions. In preparation.

EXPRESSIONS INVOLVING ADJECTIVES

We consider three types of assertions about items, classes, or events (which we designate as A, B, C, etc.) involving adjectives (which we designate as x and y for a polar pair). We also consider the synonymy of different adjectival expressions which arise from the semantic relations among polar opposites.

A is x

Examples: John is short. Elephants are big.

It has often been claimed that assertions of this type are implicitly comparative; that the meaning of the first example actually is "John is short compared to the average boy of his age," and, of the second example, "Elephants are large compared to the average of all classes of animals." Certainly, these assertions only indicate A's position on the dimension with respect to some particular superordinate class, rather than that A meets some absolute standards. Thus, it is perfectly acceptable to say, "This elephant is small" or "This turtle is fast." While the entire class of elephants is large and that of turtles is slow in comparison to various other classes, the particular items can be small or fast compared to other members of the class.

These adjectival expressions, however, can be used to assert the relation of an item to a superordinate class independent of absolute standards only in certain cases; namely, when they include the common polar adjectives from ratio scales. They cannot be used independently of absolute standards with extreme adjectives or with polar pairs from many ordinal scales. Consider, for example, the case of extreme adjectives from ratio scales. One cannot say, "This elephant is miniscule"; the adjective seems to apply only within an absolute range of sizes from which even the smallest of elephants is excluded. The same is true for extreme adjectives from interval scales. For instance, "hot" cannot be used to apply to objects normally considered cold, such as ice water, if one only means to imply that the object is less cold than the average of its superordinate class.

In addition, one cannot use ordinal scale adjectives like "happy," "sad," "pretty," "homely," etc., to compare some person, object, or event with the norm of a superordinate class without taking into consideration the absolute qualities of the items involved. For instance, in a beauty contest, it could not be asserted, "That beauty contestant is homely" just because the girl was less beautiful than the average beauty contestant. For ordinal and extreme adjectives then, expressions of the form "A is x" are not comparative as are those comparative expressions with ratio scale adjectives. To say "The water is cold," "Joan is pretty," asserts that an item meets certain absolute standards, and in that sense such adjectivals are more like assertions from noncontinuous adjective classes like "The apple is red," or "The patient is alive" than are expressions involving adjectives from ratio scales, like "This elephant is small." Thus, as one might expect, the usage of ordinal and ratio scale adjectives does differ, even though the marked adjectives from both types of scales are treated as designating the absence of the property exhibited by the unmarked adjective. That is, despite the evidence about the logical negativity of the marked adjective from both types of scales, we have here an indication that, as was mentioned above, the ordinal scale in some sense extends in both directions from a zero point and thus involves two separate properties rather than simply different amounts of the same property.

A is too x, B is not x enough

Examples: The child is not old enough to go to school. The man is too short to be a professional basketball player.

Such expressions assert that an item, class, or event is excluded from the portion of the dimension which would enable it to be a member of some category, or to be appropriate for certain purposes. Bierwisch (1967) proposes that these two types of expressions are synonymous for pairs of antonymous polar adjectives; for example, "A is too short" implies "A is not tall enough" and "A is too tall" implies "A is not short enough."

For adjectives from ratio scales, which compare an item to other items of its class independent of absolute standards, these expressions can indeed be used freely to exclude an item from any portion of the dimension; for example, while the example about the basketball player above excluded an item from only a small area at one extreme of the dimension, such expressions can also be used to exclude an item from a very large portion of the dimension, as in "The man is too short for his feet to reach the floor when he sits on a chair." However, for those adjectives which indicate that an item meets some absolute standard, these expressions *cannot* be used freely in all portions of the dimension. One cannot say, for instance, "The elephant is not miniscule enough to fit through the temple doors" (note that the expression is acceptable if the word "miniscule" is replaced by "small"). For such extreme adjectives, as well as for many polar adjectives from ordinal scales, the two forms of expression are not synonymous for pairs with opposite meaning as they are for the bipolar pairs from ratio scales; for example, "A is too cold" is not really synonymous with "A is not hot enough" nor is "A is not pretty enough" really synonymous with "A is too homely."

A is more x than B, C is not as x as D

Examples: Tom is more agile than Harry. Jane is not as smart as Mary.

The most common way to compare two items, classes, or events is the "more" form, as in "more beautiful," "more friendly," etc. For many adjectives, the compared "more" form is expressed in a single word involving the adjective plus "er" which means "more," as in "taller," "happier," etc. There is a symmetrical form of expression for explicitly comparing two items; namely, "A is less x than B." The latter is used less frequently than the synonymous form "A is not as x as B" which itself is less frequently used than the "more" form. The "A is not as x as B" form is called a "negative-equative" and derives from the equative "A is as x as B" which asserts the equality of the two items being compared. However, the negative-equative is not, in fact, the negation of the equative, since the negation would only assert that the two items are not equal without specifying direction. Rather, the "not" applies as in many statements of quantity to mean "lacking in" or "less."

It has often been claimed that the "A is more x than B" form is a short form for a sentence which says, in full, "A is x to a greater extent than B is x." One possible way to interpret such statements is to say that they are made up of two nuclear propositions joined by the relation "more than," one asserting "A is x," and the other that "B is x." Such an analysis would imply that one cannot make a comparison between two items, A and B, without attributing the quality x to both items. That this is *not* the general case is shown by the following examples: "A hippopotamus is smaller than an elephant," "A jet plane is slower than a rocket ship." According to the proposed analysis, the first comparative would presuppose the truth of the assertions "An elephant is small" and "A hippopotamus is small," and the second comparative would presuppose the truth of the assertions "A jet plane is slow" and "A rocket ship is slow." Clearly, these comparatives do not presuppose any such thing, but only that both items can be evaluated with respect to the dimension to which those adjectives belong. That is, one can compare two items using "bigger" or "smaller" as long as they are material objects, or using "faster" or "slower" as long as they move. Insofar as two items can be compared using either of the two polar adjectives, no matter where the items are located along a dimension, one can describe the same relation between the items in two alternative ways using the "more" form of comparison: "A is more x than B" or "B is more y than A."

As these examples of comparatives with ratio scale adjectives show, one must reject any general proposal that the comparative "A is more x than B" can be broken up into the two assertions "A is x" and "B is x." Nevertheless, comparatives with extreme adjectives or polar adjectives from ordinal scales which involve absolute standards do seem to presuppose that both items possess the quality x. To say "A is more obese than B" does seem to presuppose that "A is obese" and "B is obese." Similarly, to say "C is homelier than D" does seem to presuppose that "C is homely" and "D is homely."

We would tentatively propose the following: If an adjective indicates that an item meets certain absolute standards, then it is only appropriate to use it when the items meet that standard; alternatively, if an adjective can be used independent of absolute standards, then it can also be used to describe the relative positions of the two items no matter where along the dimension those items fall. Thus, for adjectives where absolute standards are involved, the two expressions "A is more x than B" and "B is more y than A" with "x" and "y" being adjectives of opposite meaning, are not synonymous; for example, "A is prettier than B" and "B is homelier than A" are not really synonymous. However, unmarked adjectives from ordinal scales can probably be used more neutrally than their marked counterparts; this may be related to the fact, noted above, that the unmarked adjective from these scales can be used nominally.

Even for comparatives involving the polar adjectives from ratio scales, the location of the items along the dimension may well have some influence on which polar adjective a speaker chooses; that is, if the items do possess the quality x to a notable extent, then, other things being equal, the speaker might use "x" in a comparative describing the relation between them. However, the location of items on the dimension is only one of the factors which determines which adjective is used. To describe a particular relation between two items using a "more" comparative, a decision about the grammatical subject will determine which polar adjective must be used. That is, when one of the items is the topic of conversation and is selected as grammatical subject, this determines the adjective. Other aspects of the speaker's focus of concern may also influence which adjective is chosen; for example, if a basketball coach has an extra player and must reject one, he may ask, "Is player A worse (shorter, slower) than player B?" Therefore, the adjective used in a "more" comparative involving ratio adjectives and some ordinal adjectives is not a sufficient clue as to the location of the items; one must also take the conversational context into account.

We would guess that there would be more reason for the listener to infer the location of the items along a dimension from a "negative-equative" than a "more" comparative. The reason is the following: The "more" form is the common way to compare two items and one would guess that a speaker would use it unless it did not allow him to express what he wanted to say. This could happen if a speaker both had reason to make one of the two items the grammatical subject, and also wanted to use a particular adjective. For example, in a discussion of the nation's poor where one's topic was migrant farm workers, one might say, "Migrant farm workers are not as poor as the urban unemployed" to avoid the adjective "rich" which would appear in the "more" comparative.

This tentative formulation of what is assumed about the items in "more" and "not as" comparatives, and in comparatives with different types of adjectives, can and should be tested empirically. For example, one can have *S*s formulate comparatives under varying conditions to determine the effects of these conditions on the type of comparative they will use; one can ask listeners to infer the location of items along a dimension given different comparative sentences in varying conversational contexts.

SYLLOGISMS

Comprehension of the Premises of the Syllogism

The ordering syllogism is a special context in which to comprehend comparatives. Syllogisms require the problem solver to determine the order of the three items described in the premises. Even if the comparative premises do provide him with information that the items are at a particular end of the dimension, such information is totally irrelevant to the reasoning task, which is to determine item order. In fact, *S*s can solve syllogisms even when the two premises involve two extreme adjectives of opposite meaning, both of which would ordinarily require items to meet certain absolute and opposing standards. For example, the premises "A is more saintly than B. C is more wicked than B" are admittedly odd, but within the context of the syllogism, people can regard them as descriptions of the relative positions of the three items on a single dimension. In fact, as we show below in the section on answering questions, there is no evidence that *S*s' representation of item order preserves any information about which adjectives were used in the premises.

The *S*s must combine information from two sources to determine the order of items in a comparative: (*a*) the adjective, which designates one of the poles, and (*b*) the type of comparative, "more" versus "not as" or "less," which indicates how the adjective is used to make the comparison. Both factors affect the difficulty of ordering items from a comparative, because for each source of information, there is both a positive and a negative form. That is, the unmarked adjective and "more" designate the presence of a property, whereas the marked adjective and "not as" or "less" designate its absence. Since the notion of absence seems to be more psychologically complex than the notion of presence, both factors affect the difficulty of a premise.

"Not as" and "less" comparatives are used less frequently than "more" comparatives, but it seems most reasonable to suppose that these forms are infrequent because they are inherently hard to interpret, rather than that they are hard to interpret because they are infrequent. In fact, we proposed in a previous paper (Huttenlocher, Higgins, Milligan, & Kauffman, 1970) that "not as" comparatives are so difficult to comprehend that *S*s tend to reorganize them into the familiar "more" form in the process of understanding them. Which "more" comparative they use ("A is not as x as B" might be reorganized either as "B is more x than A" or as "A is more y than B") depends, we hypothesized, on which item seems to them to be the topic, what type of adjective is involved, etc. When a premise has both the "not as" form and a marked adjective, there would be the greatest tendency to change it to the familiar "more" form, because such a statement, we would argue, involves a double negative. Indeed, *S*s often spontaneously describe sentences like "A is less short than B" as "double negatives."

Our *S*s frequently report that they represent information from both the adjective and the comparative form spatially when their task is to determine the order of the items, either from three-term series problems or single comparatives (as in the problem "Sam is worse than Harry. Who is better?"). The individual items themselves are generally reported as points in space with names. It is the order of these points which *S*s claim to imagine as a spatial array. DeSoto, London, and Handel (1965) have pointed out that people tend to think about adjectival dimensions spatially, with the unmarked adjective frequently thought of as being at the top and the marked adjective at the bottom. Our *S*s also report that they tend to think about "more" and "less" in terms of directional movement, with "more" being a movement in an upward direction and "less" a movement in a downward direction.

Our *S*s report that imagery is intimately tied up with determining the order of items from comparative expressions. According to *S*s' introspections, they construct imaginary spatial arrays from comparatives by imagining placing the grammatical subject and then moving the grammatical object with respect to it, using spatially represented information about the adjective and the comparison term.

DeSoto et al. (1965) have proposed that people prefer a premise to proceed from the top downward. They argue that *S*s find a syllogism easy when its premises contain adjectives which they assign to the top because then the grammatical object goes beneath the grammatical subject, allowing *S*s to proceed downward. It just so happend though, that for those adjectival dimensions where people have strong spatial preferences, it is the unmarked adjective which they prefer at the top. DeSoto et al. did not discuss the meanings of the polar adjectives per se; their proposal seems to suggest that *S*s interpret the premise before they give spatial assignments to the adjectival dimension, and that it is the way these spatial assignments are made which determines the difficulty of the problem. At least on the face of it, this proposal differs from the proposal that a syllogism is easier when its premises contain unmarked adjectives because these are less complex in meaning than marked adjectives.

As we have seen, there is evidence that the meanings of marked and unmarked adjectives differ in complexity apart from any spatial assignments these adjectives may be given. People report spatial images only when they order items along a dimension. The evidence that marked adjectives are more complex than unmarked adjectives because they include an implicit negative comes from a variety of sources, many of which do not involve spatial imagery; for example, from the way people define marked adjectives, from the way marked adjectives can be combined with other words, and from the difficulty of interpreting sentences with marked adjectives where ordering is not involved. Since the proposal that unmarked adjectives are simpler than marked adjectives covers all three cases, including ordering problems, one need not invoke spatial assignments as an explanatory principle.

To conclude that one need not postulate spatial preferences to explain the differences in difficulty between comparatives with unmarked and marked adjectives is not to say that spatial images serve no function in ordering items. Furthermore, whether or not they serve a function, the very fact that people report images in which they give consistent spatial assignments to unmarked and marked adjectives is itself a phenomenon to be explained. There are two separate issues involved: why people report spatial images at all; and given that they do so, why they consistently assign unmarked and marked adjectives upward and downward positions, respectively, and regard "more" and "less" as upward and downward movements, respectively. We will consider each of these issues in turn.

The most plausible explanation of why people report spatial images when they decode comparatives in ordering problems is that they regard ordering as a spatial operation. To order real objects along a nonspatial dimension, people assign spatial coordinates to the poles of the dimension so they can use the spatial array to preserve order information, as, for example, in arranging clothing in stores by size. If items are ordered spatially, one can easily determine what item has a particular relative position; for example, find the one which is "most x," "most y," etc. It is no more surprising that *S*s should imagine themselves to be making spatial arrays when they must order items in a syllogism than that they should imagine themselves to be moving chessmen when evaluating alternative moves in a chess game. There is, however, a question about the function of such imaginative activities in either case, an issue considered later in this paper.

The second question is why people report consistent spatial preferences rather than simply assigning spatial axes arbitrarily. We speculate that both people's spatial preferences and the fact that they encode marked adjectives in a more complex way than unmarked adjectives derive from the underlying way they think about adjectival dimensions. As we have seen, adjectival dimensions for ratio and even ordinal scales seem to designate single properties which vary in extent, with the unmarked adjective designating notable presence and the marked adjective notable absence of the same property.

Consider the possibility, then, that for both ordinal and ratio scales, people think of the extensiveness of a property in terms of mass, as if the property were a pile of some material substance. We could then explain people's spatial preferences in ordering problems—both for unmarked and marked adjectives, and for the comparative terms "more" and "less." The unmarked adjective would be toward the top because it designates the presence of a space-occupying property, and the marked adjective toward the bottom because it designates the absence of that property. "More" would be an upward movement because it designates addition, and "less" would be a downward movement because it designates subtraction. (Our pilot studies show that the unmarked adjective is represented as upward because it indicates the presence of a property, *not* because it doubles as the name of the dimension. The evidence comes from extreme pairs where neither adjective can be used neutrally, like "gorgeous–hideous" and "saintly–wicked." People regard "gorgeous" and "saintly" as spatially upward and "hideous" and "wicked" as spatially downward. In addition, syllogisms with the adjectives "gorgeous" and "saintly" in both premises are easier than those with the adjectives "hideous" and "wicked" in both premises [Higgins[4]]).

In conclusion, any spatial assignments of polar adjectives are, logically speaking, equally good for preserving information about order, as long as these spatial assignments are made consistently within a problem. The fact that *S*s exhibit consistent spatial preferences across different problems with the unmarked adjective on top, and that unmarked adjectives are simpler than marked, suggest that *S*s' assignments of the poles are related to the meanings of the adjectives. We presented a tentative hypothesis to account for these findings; namely, that adjectival dimensions are represented as properties which accumulate mass as they increase in extent. According to this hypothesis, a marked adjective, y, is difficult and is imagined downward for the same reason; that is, because it designates the absence of property x.

Combining Information from the Two Premises

An earlier paper (Huttenlocher, 1968) proposed a mechanism whereby people combine information from the two premises to determine the order of all three items. This proposal was based on a behavioral test of *S*s' introspections about how they solve these problems. The *S*s claim that they imagine the described items to be material objects to be arranged in space, and the premises to be instructions about how to arrange these imaginary items. According to their introspections, they use the first premise to create an imaginary array of the first two items. They claim to use the second premise as an instruction about how to fit the third item in among the two items from the first premise. They say that the second premise is not decoded into a separate spatial array, but rather, is used to find the position of the third item with respect to one of the items which is already positioned in the array.

Our original proposal about the construction of spatial images reported a test of this claim. We had found in other experiments that when people arrange real items according to instructions, the difficulty of arranging a mobile item with respect to one or more fixed items depends on its grammatical role in a sentence. That is, when a mobile item must be arranged with respect to fixed items, it is easiest to place when it is described as (*a*) the grammatical subject of a "more" comparative (Huttenlocher & Strauss, 1968), or as (*b*) the actor in either an active or a passive sentence with a transitive verb (Huttenlocher, Eisenberg, & Strauss, 1968). We also found that for "not as" comparatives, the difficulty of placing a real mobile item with respect to a real fixed item varies with the adjective, depending, as we argued above, on how *S*s reorganized the negative-equative in the process of understanding it.

---

[4] E. T. Higgins. A look at extreme adjectives in relation to syllogisms. In preparation.

We tested *S*s' claims about constructing imaginary spatial arrays by presenting them with syllogisms which were broken into two parts. First, *S*s were given the first premise and were asked questions about the order of the first two items. Next, they were given the second premise and were asked questions about the order of all three items. The pattern of results for the syllogisms exactly paralleled the pattern of results for the arrangement tasks. The difficulty of answering questions about the order of all three items varied with the grammatical status of the new item in the second premise, just as the difficulty of placing a real mobile item with respect to real fixed items varied with its grammatical status in the instruction. That is, a syllogism is easier when the new item in the second premise, which *S*s *imagine* to be mobile is described as (*a*) the grammatical subject in a "more" comparative, or as (*b*) the actor in either an active or passive sentence with a transitive verb (we used "leads," "is led," "trails," "is trailed"). In addition, the results for syllogisms with negative-equatives also paralleled those for arrangement tasks. That is, the results for both syllogisms and arrangement tasks varied with the adjective used, and both varied in exactly the same way.

The occurrence of a parallel between the difficulty of arrangement tasks and syllogisms is consistent with but does not prove that people imagine syllogisms to be arrangement tasks. The question of how to interpret the observed parallel requires a consideration of the reasons for the findings of the arrangement tasks. We have argued elsewhere that the arrangement studies suggest that the grammatical role of an item in the description of a transitive relation affects the difficulty of understanding how that item is related to the other item (Huttenlocher & Weiner, in press). It is for this reason that the difficulty of arranging a real item that one is trying to find out about ("new" item) with respect to an item which is fixed in an array ("old" item) varies with the grammatical role of the "new" item.

For syllogisms, there are no real items, so *S*s need not take account of the fact that one of the items in the second premise is "new" and the other is "old" when they decode the second premise. Rather, they could order the two items from the second premise, and only afterward combine this information with that from the first premise. However, the parallel between the relative difficulty of ordering the "new" item from the second premise in a syllogism and of ordering the "new" item in an arrangement task strongly suggests that *S*s do treat the third item in the second premise as "new" and the other item as "old." This is clearly consistent with their constructing spatial images. However, the grammatical role of the "new" item would affect the difficulty of understanding how it was to be ordered even if *S*s did not construct a spatial array. Thus, one could argue that it is not necessary to postulate that *S*s represent order spatially in these reasoning problems; that is, that they actually imagine adding a new item into a spatial array of fixed items. One could instead describe the process of combining information from the two premises noncommittally, without postulating the construction of spatial images. That is, one could simply postulate that *S*s order the set of three items by determining from the second premise the relation of the "new" item to the "old" item, while remembering at the same time that the "old" item is itself ordered with respect to yet another item.

To conclude that one could just as well reformulate our imagery proposal in some more abstract logical fashion would be to ignore its major strength. That is, the imagery hypothesis provides a single overall account of our experimental findings and, in addition, links *S*s' behavior to their introspections. It accounts not only for *S*s' introspections about the nature of the solution process, but also their introspections about why they make errors. That is, the spatial imagery hypothesis provides *S*s with an algorithm for obtaining answers to all soluble syllogisms, and for recognizing which syllogisms are insoluble. To attribute this strategy to our *S*s, who make many errors, we must postulate performance factors which limit their success. If *S*s do construct spatial images, they must remember *both* the names and relative positions of the first two items while decoding the second premise; so they might sometimes remember where they placed an item, but forget the names. The situation would be analogous to doing long division without pencil and paper; a person might have no problem about what procedures to follow, yet have trouble keeping track of all the necessary information. Our *S*s complained, for example, "I figured it out, but then I forgot the name of the best guy."

It is not obvious at present whether any theory which postulates imagery as a mechanism for solving problems can or cannot, in general, be reformulated in an abstract logical fashion that, nevertheless, makes the same behavioral predictions. We know from people's overt behavior that they do use various kinds of symbolic schemes to represent information; they talk, make graphs, build scientific models, etc. However, one can only infer the forms of people's *mental* representations of information insofar as these produce different behavioral consequences. If alternative forms of representation serve equivalent functions in preserving and transforming information—for example, if graphic representation is equivalent to algebraic formulation, or diagrammatic representation to verbal description—one could not identify these as distinct forms of mental representation. In that case, such questions as whether thought involves imagery, implicit speech, etc., might prove meaningless. Rather, one could only give an abstract logical specification of *S*s' strategy. The general issue of whether different types of symbolic schemes vary in their potential for representing and transforming different kinds of information has not yet been explored in depth.

A different type of behavioral evidence for imagery has been reported by Brooks (1967, 1968). He found mutual interference between visual or auditory imagery during problem solving and the intake of sensory information in the modality in which the imagery occurs. This clearly demonstrates the "psychological reality" of imagery and, in addition, provides a tool for evaluating whether such imagery has occurred in particular cases. Shaver (1970) used such an interference technique to investigate whether imagery is involved in solving ordering syllogisms. The *S*s were presented problems either visually or orally, with presentation time controlled for the two modes. (The *S*s either heard problems on tape or viewed a film where one word was presented at a time. The time between words was the same for both modes of presentation.) The *S*s did significantly better with oral rather than visual presentation. Shaver offers evidence that this finding is due to interference rather than inherent differences in the difficulty of auditory versus visual presentation. While more evidence is needed, his findings suggest tentatively that imagery is indeed involved as a mechanism for solving ordering syllogisms.

Even if there proves to be good reason to postulate that people's mental representations are isomorphic with the physically realized representations they use in solving analogous problems (graphs, maps, etc.), the major issues about human problem solving will remain to be resolved. One wants to explain how people use various symbolic schemes to preserve and transform information, regardless of whether that information is represented in thought or on paper. That is, it seems to us that the really challenging issue is raised by the fact that people are able at all to use spatial arrangements to preserve information about the order of items along nonspatial dimensions.

Answering Questions

The spatial imagery proposal does not predict differences in the difficulty of retrieving the information pertinent to alternative questions. The proposal is that, before answering any question, the problem solver creates a spatial array of all three items on the basis of which he can answer any question about the order of the three items (except, of course, when he has failed to construct the array correctly). One would expect questions to differ in difficulty depending on how hard they themselves are to understand. Thus, questions with marked adjectives should be more difficult than those with unmarked adjectives. Indeed, they are.

The difficulty of answering a question is also affected by the relation between the way the question is stated and the way the premises are stated. After examining this interaction, we believe that it is best explained by postulating that the meaning of an adjective becomes temporarily more accessible immediately after it has been used. Our reasons will become apparent from the analysis below. (Since we did not originally investigate the effect of question form on problem difficulty, we use the data from Clark, 1969.)

For both "more" and "not as" syllogisms, the first and second premises may involve either the same or different adjectives. In the former case, the adjective in the question can either match or fail to match the adjectives in the premises. In the latter case, the adjective in the question can either match the first premise and not match the second, or can match the second premise and not match the first. Regardless of whether the two premises have the same or different adjectives, Clark's data show that a question is easier to answer when the adjective in the question matches that in the second premise.

This match–mismatch effect apparently derives entirely from the second premise, since the size of the effect is equally large whether or not the adjective in the first premise *also* matches the one in the question: When the adjective is the same in both premises, the average reaction time (RT) in seconds is 5.77 when the adjective from both premises match that in the question, and 6.37 when they do not match it; when the adjective is different in the two premises, RT is 5.62 when it is only the adjective in the second premise which matches that in the question, and 6.24 when the second premise adjective does not match. Indeed, when the answer to the question is in the first premise, it is totally irrelevant whether or not that question uses the same adjective which was in the first premise. That is, when the answer is in the first premise, RT is 6.12 when the adjective in the question matches that in the first premise, and 6.00 when the adjective in the question does not match that in the first premise. (Since the effects for question answering do not differ for "more" and "not as" problems, we combined these for all our comparisons.)

This match effect is not restricted to cases where the item which is the answer is contained in the second premise. Even if the answer is in the first premise, it is still easier to answer the question if it involves the adjective used in the second premise, although not as much so as when the answer is contained in the second premise. When a question uses the same adjective as the second premise, RT is 5.42 when the answer is in the second premise and 5.97 when it is in the first, as opposed to 6.31 when the adjective in the question does not match the adjective in the second premise. (The fact that a question about an item in the second premise is easier than one about an item in the first premise may well be due to another type of recency effect.)

The fact that only the adjective in the second premise affects question difficulty suggests that the only function of the particular adjective, over and above its function in determining the order of items, is to make the meaning of the adjective in the question more accessible. That is, if a person has just determined the meaning of a particular adjective in order to comprehend the second premise, he finds it easier to determine its meaning again when that adjective appears immediately afterward in the question. If this accessibility factor is temporary, it should be sensitive to any separation between the second premise and the question. Smith, Coriell, and McMahon[5] find that the effects of match between a question and a sentence are indeed reduced if *S* must say three digits between presentation of the sentence and the question. These results suggest that there is no reason to postulate that people's mental representation of item order includes information that the items are at a particular end of the dimension. In fact, there is no reason to postulate that information about the particular adjective used is stored more than momentarily.

---

[5] K. Smith, J. Coriell, and L. McMahon. The role of active and passive voice in answering questions. In preparation.

CLARK'S PROPOSAL ABOUT HOW PEOPLE SOLVE ORDERING SYLLOGISMS

In this section, we review a paper by Clark (1969) which proposes three principles to explain how people solve syllogisms. The first principle, the primacy of functional relations, asserts that the primitive conceptual relations out of which sentences are constructed are stored in a more readily available form after comprehension than other kinds of information. Clark claims that all comparative sentences, such as "Joe is worse than Sam" can be analyzed into: (*a*) two base strings, "Joe is bad" and "Sam is bad," which he takes to be the functional relations, and (*b*) the comparison "more than"; and that *a* is more available than *b*. We rejected the claim that comparatives with ratio scale adjectives can be broken into two such propositions. We did not reject this analysis for ordinal scale adjectives or extreme adjectives. However, as we noted above, in the context of the syllogism it is only necessary to postulate *b*, not *a*. That is, it is not necessary to postulate that *S*s retain, more than momentarily, any information from the premises concerning at which end of the dimension the items are being compared.

Since Clark's first principle asserts that comparatives with "good" and "bad" have different underlying meanings, and no specification is given of the relation between these meanings, it should be impossible to solve a syllogism which has different polar adjectives in its two premises. It is only because "good" and "bad" can both be used to order items along the same dimension that it is possible to solve a problem where the two premises are "Sam is better than Harry. Joe is worse than Harry."

The second principle, the principle of lexical marking, attempts to explain the fact that comparatives involving unmarked adjectival forms are easier to interpret than those involving marked forms. It is based on the fact that unmarked adjectives can be used in either a nominal sense to indicate a dimension or in a contrastive sense to indicate one pole of the dimension, whereas marked adjectives can only be used contrastively. He points out that the nominal use can be considered lexically simpler than the contrastive use in that it indicates the dimension without indicating the pole and concludes that the unmarked adjective is easier than the marked because only it can be used nominally. First, there is no empirical evidence that nominal usage of unmarked adjectives is easier for people to understand than their contrastive usage. The factors which might lead linguists to represent contrastive usage as more complex might or might not be paralleled by factors which would lead the psychologist to conclude that contrastive usage is cognitively more complex. Second, as noted above, syllogisms with "+" pole extreme adjectives in both premises are easier than those with "−" pole extreme adjectives in both premises even though the "+" pole extreme adjectives *cannot* be used nominally. Finally, if the differences in difficulty between expressions involving marked and unmarked adjectives result from the unmarked adjective being used in a nominal sense, then expressions which involve the contrastive senses of unmarked and marked adjectives should be equal. However, Sherman found significant differences between marked and unmarked adjectives where both were used contrastively; for example, "The man is not sad" is more difficult than "The man is not happy."

To pursue his argument, Clark adopts a peculiar position which he mistakenly takes to derive from linguistic theory; namely, that comparatives with unmarked adjectives involve nominal use and comparatives with marked adjectives involve contrastive usage. This would suggest that the unmarked form could be used to make comparisons anywhere within the scale but that the marked form could only be used near one end of the scale; for example, that "taller" or "faster" could be used to compare items throughout the dimensions of height or speed, but that "shorter" or "slower" could only be used to compare items that were "short" or "slow." This is a false claim; recall our example above, "A jet plane is slower than a rocket ship." The reader can easily verify for himself that, for adjectives from ratio scales, there is no such asymmetry between the breadth of applicability of marked and unmarked forms in "more" comparatives.

In any case, this second principle conflicts with the analysis of the comparative construction in the first principle. Recall that Clark claims that these include two underlying propositions, each clearly involving contrastive usage, regardless of whether the adjective is marked or unmarked. Clark never resolves this contradiction by retracting his first principle for the case of unmarked adjectives and suggesting some other analysis.

Given this contradiction, it is difficult to consider the predictions to which the analysis might lead. Since the second principle only provides a rationale for the finding that comparatives are easier with unmarked rather than marked adjectives, let us consider the predictions from the first principle. This principle has two parts: that the comparative is analyzed into two simple propositions and that the information in each of the two is more accessible than the comparative information about the relation between the two items. The second part, the matter of relative accessibility, is never used to make any predictions. The first part, that *S*s do analyze comparatives into such underlying assertions, is tested in conjunction with his third principle, "congruence."

The principle of congruence asserts that the form of a question must match the form of stored information concerning the correct answer in order for *S* to be able to answer. Clark does find that in the case of single comparative sentences it is easier to answer "better" questions for comparatives involving "good" (both "A is better than B" and "B is not as good as A") and "worse" questions for comparatives involving "bad." This fact is taken to verify both the first and third principles, namely, that the underlying form of "good" comparatives must include "good" and that congruence between statement and question is necessary. As was pointed out above, when we proposed a "temporary accessibility" factor, what is important is the relation between the adjective in the question and the adjective in the second premise, not whether the adjective of the question matches the adjective of the premise which contains the answer. In addition, this congruency principle would not predict that match effects between a question and a premise would be reduced by the interpolation of digits as found by Smith et al. (see Footnote 5).

These three principles concern the storage and retrieval of information from individual comparative sentences. They do not involve any proposal about how information from these individual comparative sentences is combined to yield the order of all three items. In order to explain effects due to combining information from two sentences—notably, the fact that reordering a particular pair of premises has marked effects on problem difficulty—Clark invokes the notion of "compression," which is in no way related to his other three principles. This notion is introduced as if it were an empirical generalization which is of no theoretical significance. It is nevertheless made to bear the major burden in predicting differences between alternative forms of the three-term series problems.

"Compression" refers to a postulated tendency to retain only one of two items from the first premise; for example, from the premise "John is better than Bill," the claim is that *S*s retain only "John is the better one." When the second premise is either "Sam is better than John" or "John is worse than Sam," the name retained from the first premise is mentioned in the second. Although Clark does not indicate how *S*s combine this information about John with the information about John and the third person in the second premise, it seems possible that such a mechanism might be worked out; but, even then, only one of the two possible questions could be answered. For instance, in our example above, the question "Who is worst" is unanswerable, as information about Bill has been lost. When the second premise is either "Sam is worse than Bill" or "Bill is better than Sam," however, the item *S* retains from the first premise is not even mentioned in the second. In this case, the information about John from the first premise bears no relation to the information about Sam and Bill from the second premise. Clark (1969) is aware of this problem and proposed that

> in this case *S* must try to recover the whole first proposition or back-track with the information that "Bill is the better one" to apply it to the first proposition or revert to some other time-consuming strategy [p. 397].

Since Clark does not tell us what this time-consuming strategy might be, he cannot be said to have provided a mechanism by which *S*s can solve the problems in these cases. Furthermore, the claim that these problems require a time-consuming strategy is contradicted by Clark's data as well as that of other investigators. The problems which are easiest overall do *not* mention in the second premise the item Clark claims that *S*s retain from the first; namely, problems with the following premises: "A is better than B; C is worse than B" or "C is worse than B; A is better than B."

In addition to the above criticisms of Clark's principles, there is further empirical evidence indicating that one or more of his principles must be incorrect, since together they make predictions which are contradicted by Clark's own data.

As an example, consider the two forms (*a*) "A is better than B, B is better than C; who is best?," and (*b*) "C is worse than B, B is worse than A; who is worst?," and compare them with the forms (*c*) "B is worse than A, B is better than C; who is best?" and (*d*) "B is better than C, B is worse than A; who is worst?" We pick this comparison because, within the context of Clark's principles, one cannot possibly explain why *a* and *b*, where the question is congruent, should not be easier than *c* and *d*, where the question is incongruent. First, "B" is described as both "bad+" and "good+," in problems *c* and *d*, which, as we pointed out above, should render these problems insoluble. Second, if *S*s compress the first premise, then *a* and *b* must be easier than *c* and *d* because, for the former, the item retained from the first premise (A or C, respectively) is the answer, whereas, for the latter, the item retained (B in both cases) is *not* the answer. Third, while *a* and *b* versus *c* and *d* do differ in terms of the principle of lexical marking, this factor should have no overall effect on difficulty because *a* has two unmarked adjectives and *b* has two marked adjectives whereas *c* and *d* each have one premise with a marked and one with an unmarked adjective. Certainly then, *c* and *d* should have longer latencies than *a* and *b* if Clark's approach is correct. This is not the case. (The latencies in seconds from Clark's data are as follows: $c = 5.00$ and $d = 5.45$ versus $a = 5.42$ and $b = 6.53$.) The reader can find similar contradictions in the negative-equative problems.

CONCLUDING REMARKS

In this paper, we have considered how people process various types of adjectives and comparative sentences; how they combine information from two comparative expressions to determine the order of all three items; and how they answer questions posed in the problems. Many of the issues we have considered cannot be fully resolved to our satisfaction at present, particularly the issue of whether imagery can be conclusively demonstrated to be superior to more abstract logical formulations in explaining the behavioral data related to combining information from the two syllogistic premises. In any case, the obvious difficulty of resolving this question raises the interesting and more general issue of whether different types of symbolic schemes vary in their potential for representing and transforming different kinds of information.

REFERENCES

BIERWISCH, M. Some semantic universals of German adjectivals. *Foundations of Language*, 1967, **3**, 1–36.

BROOKS, L. R. The suppression of visualization by reading. *Quarterly Journal of Experimental Psychology*, 1967, **19**, 289–299.

BROOKS, L. R. Spatial and verbal components of recall. *Canadian Journal of Psychology*, 1968, **22**, 349–368.

CLARK, H. R. Linguistic processes in deductive reasoning. *Psychological Review*, 1969, **76**, 387–404.

CLARK, H. R., & CARD, S. K. Role of semantics in remembering comparative sentences. *Journal of Experimental Psychology*, 1969, **82**, 545–553.

DESOTO, C., LONDON, M., & HANDEL, S. Social reasoning and spatial paralogic. *Journal of Personality and Social Psychology*, 1965, **2**, 513–521.

HUTTENLOCHER, J. Constructing spatial images: A strategy in reasoning. *Psychological Review*, 1968, **75**, 550–560.

HUTTENLOCHER, J., EISENBERG, K., & STRAUSS, S. Comprehension: Relation between perceived actor and logical subject. *Journal of Verbal Learning and Verbal Behavior*, 1968, **7**, 527–530.

HUTTENLOCHER, J., HIGGINS, E. T., MILLIGAN, C., & KAUFFMAN, B. The mystery of the "negative equative" construction. *Journal of Verbal Learning and Verbal Behavior*, 1970, **9**, 334–341.

HUTTENLOCHER, J., & STRAUSS, S. Comprehension and a statement's relation to the situation it describes. *Journal of Verbal Learning and Verbal Behavior*, 1968, **7**, 300–304.

HUTTENLOCHER, J., & WEINER, S. L. Comprehension of instructions in varying contexts. *Cognitive Psychology*, in press.

KLIMA, E. S. Negation in English. In J. Fodor & J. Katz (Eds.), *The structure of language*. Englewood-Cliffs, N. J.: Prentice-Hall, 1964.

MANN, J. W. Defining the unfavorable by denial. *Journal of Verbal Learning and Verbal Behavior*, 1968, **7**, 760–766.

SHAVER, P. R. Interference with spatial imagery during problem solving. Unpublished doctoral dissertation, University of Michigan, 1970.

SHERMAN, M. Some effects of negation and adjectival marking on sentence comprehension. Unpublished doctoral dissertation, Harvard University, 1969.

THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words*. New York: Columbia University, Teachers College, Bureau of Publications, 1944.

ZIMMER, K. E. Affixal negation in English and other languages: An investigation of restricted productivity. *Word*, 1964, **20** (Monogr. Suppl. 5).

(Received January 8, 1971; revision received June 22, 1971)

Psychological Review
1971, Vol. 78, No. 6, 505–514

MORE ABOUT "ADJECTIVES, COMPARATIVES, AND SYLLOGISMS":

A REPLY TO HUTTENLOCHER AND HIGGINS[1]

HERBERT H. CLARK[2]

Stanford University

In attempting a *rapprochement* of J. Huttenlocher's imagery theory and H. H. Clark's linguistic theory of reasoning, J. Huttenlocher and E. T. Higgins first accepted the imagery theory as valid and second, criticized and revised the linguistic theory. This reply discusses available evidence, not considered by Huttenlocher and Higgins, which shows instead (*a*) that the imagery theory is empirically wrong and (*b*) that each criticism and revision Huttenlocher and Higgins made of the linguistic theory fails on empirical grounds. In addition, this reply points out several fundamental misrepresentations Huttenlocher and Higgins made of the linguistic theory as well as their incorrect use of the terms "marked" and "unmarked." It is argued, in sum, that the available evidence invalidates those authors' "more complete explanation of how people solve ordering syllogisms."

Huttenlocher and Higgins (1971) attempted a *rapprochement* between the reasoning theories of Clark (1969a) and Huttenlocher (1968), hereafter the linguistic and imagery theories, respectively, both of which were designed to account for how people solve three-term series problems (e.g., "John is better than Pete, John is worse than Sam; who is best?"). In their attempt, they accepted the imagery theory without any fundamental change and discussed the linguistic theory in detail, rejecting each of its proposals and devising others where needed. But in their paper, they failed to consider a sizable body of available evidence which shows that (*a*) the imagery theory, far from being totally acceptable, is empirically wrong, (*b*) their rejections of the linguistic theory's principles of congruence and lexical marking fail empirically, (*c*) their rejection of the principle of functional relations fails on both psychological and linguistic grounds, and (*d*) their notion of "temporary accessibility," replacing the principle of congruence, is disconfirmed by the data. In addition, Huttenlocher and Higgins used the terms "marked" and "unmarked" incorrectly throughout most of their paper. These deficiencies in Huttenlocher and Higgins' paper are serious, for, if real, they overturn each of its major points and totally invalidate their attempt at a "more complete explanation of how people solve ordering syllogisms." Thus, it is important to consider each of these deficiencies in detail.

DISCONFIRMATION OF THE IMAGERY THEORY

In the imagery theory, the subject is seen as placing the three terms of a problem into a mental array. For the premise combination "Abel is better than Baker, Charlie is worse than Baker," for example, the subject first places Abel and Baker of the first premise into the array, then he places the "new" term of the second premise (Charlie) into the array with respect to the "old" term of the second premise (Baker). It is the grammatical function of the "new" term in the second premise that is crucial. Whenever it is the subject of the second premise (as Charlie is, above), placement is easy; whenever it is not, placement is difficult. This prediction derives from actual placement tasks in which people find it easier to place a "new" object into a physical array when it is the subject of a comparative sentence than

---

[1] The preparation of this reply was supported in part by United States Public Health Service Research Grant MH-20021 from the National Institute of Mental Health. I am indebted to P. Carpenter, E. V. Clark, M. Just, C. E. Offir, E. E. Smith, and D. G. Tieman for comments.

[2] Requests for reprints should be sent to Herbert H. Clark, Department of Psychology, Stanford University, Stanford, California 94305.

TABLE 1

PERCENT ERRORS FROM CLARK (1969b) IN "DETERMINATE" PROBLEMS FAVORED AND DISFAVORED BY THE IMAGERY THEORY

Problems	Favored problems	Disfavored problems	Disfavored −favored
Comparative			
Premises with same adjective	17	34	+ 17**
Premises with different adjectives	10	32	+ 22**
Total	13	33	+ 20**
Negative-Equative			
Premises with same adjective	33	24	− 9*
Premises with different adjectives	40	28	− 12**
Total	36	26	− 10**

Note.—Each nontotal percentage is based on 400 responses.
* $p < .01$.
** $p < .001$.

when it is not. This theory, however, is disconfirmed by two independent pieces of evidence, both derived from data not considered by Huttenlocher and Higgins.

1. In Clark (1969a), the imagery theory was put to an explicit test by making use of problems containing only negative-equative premises (e.g., "Baker isn't as good as Abel"). This test depended on only one assumption: *Assumption 1:* People find it easier in an actual placement task to place a "new" item into an array whenever it is the *subject* of a negative-equative sentence. Given the imagery theory plus Assumption 1, it was possible to classify every three-term series problem in Clark (1969a) as either favored by the theory (because its "new" term is the subject of the second premise) or as disfavored (because its "new" term is in the predicate of the second premise). The error data from Clark (1969b), summarized in Table 1, show that whereas the imagery theory is correct in its predictions for problems containing comparative premises, it is consistently *incorrect* for problems containing negative-equatives. Therefore, if Assumption 1 is correct, the imagery theory is disconfirmed by these data as well as by parallel data in Clark (1969a) and Huttenlocher, Higgins, Milligan, and Kauffman (1970).

To explain these negative-equative data, however, Huttenlocher et al. more recently contended, instead, that Assumption 2 is the proper one to make: *Assumption 2:* People find it easier in actual placement tasks to place a "new" item into an array whenever it is the *second* term of certain types of negative-equative sentences (e.g., those with "good" or "bad"). As evidence, they pointed to two tasks. In Task 1, subjects solved three-term series problems while explicitly manipulating figures on a felt board; in Task 2, other subjects solved the same problems in their heads. The results of Tasks 1 and 2 were parallel. From this, Huttenlocher et al. argued: (*a*) Task 1 revealed the difficulties subjects had in manipulating objects overtly; therefore (*b*) the results of Task 2 were parallel because the Task 2 subjects were also manipulating objects in arrays, although covertly. For all this to work out, Huttenlocher et al. had to make Assumption 2 based on the supposition that subjects implicitly "convert" premises like "Abel isn't as bad as Baker" into "Baker is worse than Abel" in such reasoning tasks. Logically, however, Huttenlocher et al. could equally well have argued for the reverse: (*a*) Task 2 revealed the linguistic difficulties subjects had in taking in and using the language of these problems; therefore (*b*) the results of Task 1 were parallel because the Task 1 subjects were also prey to these linguistic difficulties. Thus, because all difficulties in both tasks could be linguistic in origin, Huttenlocher et al.'s evidence cannot distinguish between Assumption 1 and Assumption 2. The resolution of this issue therefore requires an independent and more direct test of Assumptions 1 and 2.

Evidence reported in Clark (1970a, p. 300) provides just such a test, and it shows that Assumption 1 is correct and Assumption 2 is incorrect. The evidence is that adults find it easier in placement tasks to place a "new" item into an array when it is expressed as the grammatical subject in *both* the comparative sentence *and* the negative-equative sentence. More recent unpublished experiments[3] have replicated this consistent result in considerable detail, showing how subjects could not have been making the "conversions" Huttenlocher et al. assumed and explaining why Huttenlocher et al.'s Task 1 and Task 2 had parallel results. In other words, the original disconfirmation of the imagery theory in Clark (1969a) still holds, since it was based on the correct Assumption 1.

---

[3] H. H. Clark. Difficulties people have in answering the question "Where is it?" In preparation.

2. Other data from Clark (1969b), on the so-called "indeterminate" problems, provide an independent disconfirmation of the imagery theory without the use of Assumption 1 or Assumption 2. In "indeterminate" problems, the two premises (e.g., "Abel is better than Baker, Alan is better than Baker") describe the three terms in an indeterminate order. For this particular pair of premises, one cannot tell whether Abel or Alan is better, so when asked "Who is best?" one must answer "cannot tell," although when asked "Who is worst?" one is able to answer "Baker." The imagery theory applies to these problems too, for, as Huttenlocker and Higgins state, "the spatial imagery hypothesis provides *S*s with an algorithm for obtaining answers to all soluble syllogisms, and for recognizing which syllogisms are insoluble [pp. 498–499]." And like the "determinate" problems, each "indeterminate" problem can be classified as either favored or disfavored by the theory.

As the data summarized in Table 2 indicate, there is no overall difference in errors between favored and disfavored problems. And since no differences emerge even for comparative and negative-equative problems taken separately, the imagery theory is disconfirmed regardless of whether one makes Assumption 1, as was done in Table 2, or Assumption 2. Yet if one does make Assumption 2 along with Huttenlocher et al., then the comparative and negative-equative problems with the same adjective in both premises show a significant difference ($\chi^2 = 5.00, df = 1, p < .05$) in the *wrong* direction. The 0% difference between favored and disfavored problems in Table 2 should be compared to the large 20% difference in the comparative problems of Table 1, the difference the imagery theory was originally designed to account for. The linguistic theory, in contrast, accounts for all the major differences in these "indeterminate" problems.

TABLE 2

PERCENT ERRORS FROM CLARK (1969b) IN "INDETERMINATE" PROBLEMS FAVORED AND DISFAVORED BY THE IMAGERY THEORY

Problems	Favored problems	Disfavored problems	Disfavored −favored
Comparative			
Premises with same adjective	13	10	− 3*
Premises with different adjectives	37	40	+ 3
Total	25	25	0
Negative-Equative			
Premises with same adjective	19	24	+ 5*
Premises with different adjectives	53	48	− 5
Total	36	36	0

Note.—Each nontotal percentage is based on 400 responses.
* $p < .10$.

To summarize briefly, two independent pieces of available evidence, not considered by Huttenlocher and Higgins, show that the imagery theory is empirically wrong.

ON HUTTENLOCHER AND HIGGINS' CRITICISMS OF THE LINGUISTIC THEORY

In considering the linguistic theory, Huttenlocher and Higgins rejected the principles of congruence, lexical marking, and functional relations as given in Clark (1969a) and proposed certain fundamental revisions. This section details how each of their criticisms and revisions fails on empirical grounds and, furthermore, how they erred in their use of certain linguistic terms and linguistic evidence.

The Principle of Congruence

This principle posits, roughly, that a question should be easier to answer whenever the question and the premise containing the answer are congruent in their underlying representations. After examining certain solution times in Clark (1969a), Huttenlocher and Higgins rejected this principle, since they found that the question *appeared* to be easier only when it was congruent with the second premise. So, as an alternative to congruence, they suggested that "the meaning of an adjective becomes temporarily more accessible immediately after it has been used [p. 500]."

TABLE 3

Two Exemplary "Indeterminate" Problems and Their Principal Errors

Problem	Erroneous answer	Percentage
VII. If Abel is better than Baker, And Baker is worse than Alan, Then who is best?	Abel (congruent)	26
	Alan (incongruent)	6
	Baker (neither)	2
VII′. If Abel isn't as bad as Baker, And Baker isn't as good as Alan, Then who is best?	Abel (incongruent)	12
	Alan (congruent)	30
	Baker (neither)	2

Note.—Each problem was attempted by 100 subjects.

In relying on only the solution times in Clark (1969a) for evidence, however, Huttenlocher and Higgins rejected congruence and accepted their own "accessibility" notion on a single unreplicated result. While they made detailed use of the solution times, which were drawn from 13 subjects with two observations per subject per problem, they bypassed the more precise error data collected on these same problems from 100 subjects in Clark (1969b). The latter data give complete support to congruence and no support to "accessibility." First consider only the "determinate" problems, the same problems on which Huttenlocher and Higgins based their rejection of congruence. When both premises contain the same adjective, problems favored and disfavored by congruence are also favored and disfavored by "accessibility," and these produced 24% and 30% errors, respectively. This result parallels the solution times. But when the two premises contain different adjectives, congruence and "accessibility" make different predictions. The problems favored and disfavored by congruence produced 19% and 36% errors, respectively, in parallel with the solution times and in agreement with the data from all previous experiments (as reviewed in Clark, 1969a). But the problems favored and disfavored by "accessibility" produced 28% and 27% errors, respectively, a result slightly in the *wrong* direction. This difference is not parallel to the solution times in Clark (1969a) and suggests that the difference favoring "accessibility" in the solution times was a chance effect. This conclusion is confirmed by an examination of the "indeterminate" problems. "Accessibility" was favored 16% to 16% in problems with the same adjective in both premises and 44% to 45% in problems with different adjectives in the two premises, both differences being nonsignificant. In contrast, congruence was supported 18% to 23% (a significant difference) in the first type of problem, whereas no such direct classification of problems is possible with the latter type of problem (but see below for other predictions). To summarize, where the predictions of congruence and "accessibility" are different, congruence is supported by (*a*) the latencies on "determinate" problems, (*b*) the error data on "determinate" problems, and (*c*) the error data on "indeterminate" problems, with no counterevidence. "Accessibility," on the other hand, is supported only by the unreplicated difference Huttenlocher and Higgins noted in the solution times of "determinate" problems. Congruence is well established by these data. "Accessibility" is not.

In their rejection of congruence, Huttenlocher and Higgins also failed to consider perhaps the most striking evidence supporting congruence—the specific errors the subjects in Clark (1969b) made on "indeterminate" problems with different adjectives in the two premises. Consider the two illustrative problems in Table 3. Although the correct answer to both problems is "cannot tell," subjects frequently erred, most often giving one of the two "better" terms (Abel or Alan) as answers. Of these two terms, however, subjects overwhelmingly chose the term from the premise congruent with the question, Abel for Problem VII and Alan for Problem VII′. Over all eight such "cannot tell" problems, the choice was 26% to 9% for the congruent term, a result predicted by congruence but not by "accessibility." This result illustrates one of the strengths of the linguistic theory: it predicts not only *how many* errors subjects should make, but also, in many instances, *which* errors. The imagery theory predicts only how many, thereby failing to account for these data.

Congruence, moreover, has been demonstrated to be a very general cognitive principle in other studies. It has been shown to apply to verification and question-answering in tasks using negatives (Chase & Clark, in press; Clark, 1970b, in press; Trabasso, Rollins, & Shaughnessy, 1971), locatives of various sorts (Clark, in press), actives and passives (Clark, in press), as well as comparatives. Thus, to reject congruence and replace it by "accessibility" generally, Huttenlocher and Higgins would have to contend with this much wider range of phenomena. Yet the latter support congruence and will not admit an explanation based simply on the "temporary accessibility" of a word's meaning. One must conclude that the single unreplicated result Huttenlocher and Higgins quote is insufficient to support their case.

Principle of Lexical Marking

A major part of Huttenlocher and Higgins' paper consists of a detailed recapitulation of the linguistic evidence for marked and unmarked adjectives presented in Clark (1969a) plus a discussion of some other interesting evidence on adjectival scales. The main goal of all this was to consider two proposals for the advantage of unmarked over marked adjectives (e.g., "good" over "bad") when used in three-term series problems: *Proposal 1:* An unmarked adjective used in its "nominal" sense is easier to comprehend than an unmarked or marked adjective used in its "contrastive" sense. *Proposal 2:* Certain adjectives are the "negatives" of either their unmarked or their positive counterparts, and for this reason, the former adjectives are more difficult to comprehend. Huttenlocher and Higgins rejected Proposal 1 from Clark (1969a)—although on no negative evidence—and suggested Proposal 2 in its place. Proposal 2 is reasonable and has been previously made in two recent revisions of the linguistic theory (Clark, 1970b, in press). On the other hand, the rejection of Proposal 1 is wrong.

Huttenlocher and Higgins rejected Proposal 1 on the grounds that "there is no empirical evidence that nominal usage of unmarked adjectives is easier for people to understand than their contrastive usage [p. 501]." But they did not consider the solution times, collected by Jones (1970), on three-term series problems in which the adjective of both premises was "thicker" or "thinner," or "fatter" or "thinner." As Jones pointed out, "thick–thin" is an unmarked–marked pair, whereas "fat–thin" is not, so "thick" can be used in a nominal sense in comparatives, whereas "fat" cannot. But "fat–thin," according to Huttenlocher and Higgins, is a positive–negative pair. Proposal 1 alone predicts an advantage of "thicker" over "thinner" in Jones' data, but no advantage of "fatter" over "thinner." Proposal 2 alone predicts the same advantage of "fatter" over "thinner" as for "thicker" over "thinner," since "thinner" is "negative" in both pairs. Jones' results show a significant 4.5-second advantage of "thicker" problems over "thinner" problems, but only a nonsignificant 1.9-second advantage of "fatter" problems over "thinner" problems. This constitutes support for Proposal 1 independent of Proposal 2, contrary to Huttenlocher and Higgins' claim of "no empirical evidence." Jones' analogous results on "heavier–lighter" (an unmarked–marked pair) versus "darker–lighter" (not an unmarked–marked pair) also constitute support for Proposal 1, no matter whether one considers the "positve" member of the latter pair to be "dark" or "light."

The conclusion drawn in Clark (in press), therefore, was that the attested advantage of unmarked over marked adjectives in the pairs "higher–lower," "taller–shorter," "longer–shorter," "farther–nearer," "thicker–thinner," "deeper–shallower," "bigger–wee-er," "more–less," "faster–slower," "warmer–colder," "better–worse," "happier–sadder," and "older–younger" (cf. Clark, 1969a) could be accounted for by Proposal 1 or Proposal 2 or both in conjunction. At this point, there is evidence for both proposals and no evidence to reject either one. Furthermore, Huttenlocher and Higgins' unqualified use of the word "negative" in discussing Proposal 2 leaves the impression that marked adjectives like "short" should behave psychologically like

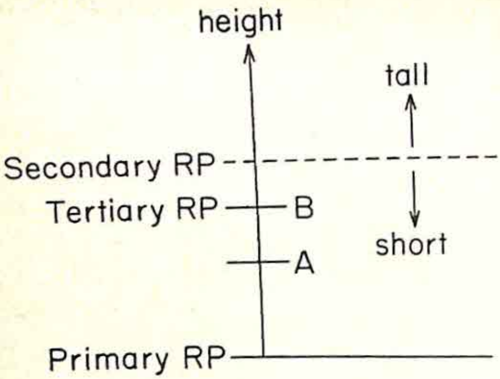


FIG. 1. A schematic diagram of the primary, secondary, and tertiary reference points (RPs) on the scale of height for the comparative "A is shorter than B."

the negatives "not, "no," none," etc. This is not the case. In timed verification tasks, the pattern of reaction times to true and false sentences containing marked adjectives is like that for positive sentences, *not* like that for sentences with "explicit negatives," even though marked adjectives without -er take longer to verify than unmarked ones (cf. Clark, 1970b; Just & Carpenter, 1971).

An additional serious problem in Huttenlocher and Higgins' discussion of Proposals 1 and 2 is that they continually misapplied the terms "unmarked" and "marked," two terms central to their paper. They first defined "unmarked" adjectives as those (like "long") that can take on a neutral sense in questions (How long is the board?), in the scale name (length), or in quantified expressions (two feet long); this definition, also found in Clark (1969a), agrees with the linguistic application of these terms (Greenberg, 1966; Lyons, 1968, p. 466). But then, without indicating a change in definition, Huttenlocher and Higgins applied the term "unmarked" to any term they thought was "positive." In their Table 1, for example, they refer to "hot–cold," "noisy–quiet," "rich–poor," "dirty–clean," "hard–soft," "fat–thin," and "smart–stupid" as unmarked–marked pairs, even though none of these fulfills the proper linguistic conditions. How they actually did make their classification is not clear. They might have required that the alleged "unmarked" adjective be positive, where "positive" is defined as an excess on the scale underlying the pair of adjectives. So "rich" would be positive since it denotes an excess of "amount of money." But even this criterion fails, for example, on "hard–soft." Does "hard" refer to an excess on "amount of hardness" or to a lack on "amount of softness?" Their choice of the first possibility seems arbitrary and unmotivated. It was their confusion of "unmarked" and "positive" perhaps that led them to overlook Jones' evidence, discussed above, on "thicker–thinner" versus "fatter–thinner" and its relevance to Proposals 1 and 2.

Principle of Functional Relations

Huttenlocher and Higgins question this principle mainly by questioning the analysis of the comparative given in Clark (1969a). This analysis, presented explicitly as simply a brief paraphase of well-known linguistic analyses by Lees (1961), Smith (1961), Chomsky (1965), Huddleston (1967), and Doherty and Schwartz (1967), noted that "John is better than Pete" is derived, by a series of linguistic rules, from the two "base strings" "John is good" and "Pete is good" or, as expressed in Clark (in press), "from something like that which underlies *John is so good* and *Pete is so good*, where *John is so good* means something like 'John is good to an unspecified degree.'" Huttenlocher and Higgins' attack on this discussion is based on several fundamental misunderstandings of this linguistic analysis.

The main problem is that Huttenlocher and Higgins mistakenly interpret the linguistic analysis as claiming that "John is better than Pete" implies that both John and Pete are good in an absolute sense. Their first remark, "Clark claims that these [comparative constructions] include two underlying propositions, each clearly involving contrastive usage [e.g., that found in "John is good"], regardless of whether the adjective is marked or unmarked [p. 502]." But this is explicitly contradicted in the following quotations from Clark (1969a): "It will be argued that the *goods* [underlying "John is better than

Pete"] can be either nominal or contrastive, but the *bads* [underlying "Pete is worse than John"] must be contrastive [p. 391]." And: "The *intelligent* underlying [This idiot is more intelligent than the idiot] must therefore be nominal, so that the base string *this idiot is intelligent* really only means 'this idiot has measurable intelligence' [p. 391]." Later, Huttenlocher and Higgins correctly pointed out that "Jets are slower than rockets" does not presuppose the truth of the assertions Jets are slow and Rockets are slow, but they incorrectly claimed that this contradicts the linguistic analysis of the comparative. In making this claim, they have assumed that the linguistic analysis disallows the use of *fast* and *slow* in their "relative" senses in comparatives. This assumption is false.

To understand this point, consider "A is shorter than B" and the three zero points, or reference points, implicit in it. As illustrated schematically in Figure 1, the primary reference point is the zero point for the whole scale of "height" and defines the zero for the "nominal" *tall* (as in six feet tall). The secondary reference point is the temporary standard or middle point to which objects are compared; objects above the standard are "tall," and those below it are "short." Thus, the secondary reference point is the zero for both "contrastive" *tall* (from there upward) and "contrastive" *short* (from there downward). Finally, the tertiary reference point is the standard chosen for the particular comparison "A is shorter than B," where A is being described relative to B. In this schema, the secondary reference point can range almost anywhere on the scale, for "tall" and "short" are relative and not absolute adjectives. One can say "John is very tall for a dwarf, but very short for a giant," successively using two quite disparate secondary reference points. In terms of this schema, the linguistic theory makes Claim 1, and the imagery theory, Claim 2: *Claim 1:* The choice of "taller" for comparing A and B presupposes that the secondary reference point is at least as low as A and B; the choice of "shorter" presupposes the opposite. *Claim 2:* The choice of "taller" or "shorter" for comparing A and B makes no presuppositions about the positions of A and B relative to the secondary reference point. According to Claim 1, "A is shorter than B" presupposes something like "A and B are short to an unspecified, perhaps null, extent relative to the secondary reference point." Thus, "A is shorter than B" does presuppose the shortness of A and B, but only relative to the secondary reference point, whose position is perhaps very high in absolute terms.

The evidence Huttenlocher and Higgins quote against Claim 1—that "Jets are slower than rockets" does not presuppose that jets and rockets are slow—is therefore defective. This sentence presupposes simply that jets and rockets are slow relative to some very high standard speed, the secondary reference point. This is possible because the secondary reference point can be placed almost anywhere, as in "Jets are slow for modern aerial weaponry." When the secondary and tertiary reference points are identical, in fact, one can say "Jets are slow relative to rockets." The last two sentences both use a "contrastive" *slow*, as required by the linguistic theory, but neither implies an absolute "Jets are slow," contrary to Huttenlocher and Higgins' interpretation of the linguistic analysis. In short, the evidence they cite is irrelevant to Claim 1.

But there is also positive evidence for Claim 1 and against Claim 2. Claim 1 is that (*a*) "Jets are slower than rockets" and (*b*) "Rockets are faster than jets," when referring to the same situation, have different secondary reference points, one high and one low, respectively. The linguistic theory claims that this difference should affect reasoning in accordance with the principle of congruence. The imagery theory denies Claim 1 and any possible effect of this difference on combining information from premises, for, as Huttenlocher and Higgins allege, "there is no evidence that subjects' representation of item order preserves any information about which adjectives were used in the premises." But all the support for the principle of congruence detailed above contradicts this allegation, supporting Claim 1 and disconfirming Claim 2. The allegation just quoted, however, is false in an even more obvious way, for some problems *require* the use of the "information about which adjectives were used in the premises." To solve the problem "If Pete is poorer than Dick, and John is rich, then who is poorest?" (answer: Pete), subjects are forced to use the presuppositional information about Pete and Dick's relative poorness, information that Huttenlocher and Higgins agree is found in the first premise. This problem, which can be solved, cannot be solved any other way. It is a small step from here to assume that the same information facilitates the solution of "If Pete is poorer than Dick, and John is richer than Dick, then who is poorest?" (answer: Pete), where the same presuppositional information is congruent with the question. But the presuppositions hinder the solution of "If Pete is poorer than Dick, and Pete is richer than John, then who is the poorest?" (answer: John), where the presuppositional information that John is rich is incongruent with the question. This, in essence, is the position of the linguistic theory, and it agrees with all the data.

Although Huttenlocher and Higgins deny any difference in meaning between (*a*) "Rockets are faster than jets" and (*b*) "Jets are slower than rockets," and other similar pairs, a closer look shows that they cannot be right. The presuppositions of *a* and *b*, though subtle, become highlighted when *a* and *b* are altered in minor ways. Note, first, that (*c*) "Rockets are no faster than jets" and the slightly odd (*d*) "Jets are no slower than rockets" both mean "Rockets and jets have the same speed," yet they are not semantically equivalent. The addition of the simple negative emphasizes relative fastness in *c* and relative slowness in *d*. To see this, contrast the sentence "Jets and rockets are very fast, but jets are no faster than rockets" against the unacceptable "*Jets and rockets are very fast, but jets are no slower than rockets." The *evens* in (*e*) "Rockets are even faster than jets" and (*f*) "Jets are even slower than rockets" have a similar effect, just as they do with the noncomparative terms in (*g*) "Rockets are fast even relative to jets" and (*h*) "Jets are slow even relative to rockets." Also, many people find the problem (*i*) "If Dick runs slower than Pete, and John runs fast, then who is slowest?" determinate, just as they should if they keep the same secondary reference point in the two premises. But they do not find the problem (*j*) "If Pete runs faster than Dick, and John runs fast, then who is slowest?" determinate, again as the reference point analysis would predict. In contrast, Huttenlocher and Higgins would claim that Problems *i* and *j* are equivalent in their indeterminacy.

Huttenlocher and Higgins made at least two other false claims in their discussion of the principle of functional relations. First, they issued the following challenge: "The reader can easily verify for himself that, for adjectives from ratio scales, there is no such asymmetry between the breadth of applicability of marked and unmarked forms in 'more' comparatives [p. 502]." Yet there is an obvious set of counterexamples. Note that "The desk is shorter than the door is wide" is acceptable, but "*The door is wider than the desk is short" is not. In sentences like these, the second adjective must be unmarked whenever it is on a ratio scale. Here, just as with adjectives from "ordinal" scales (cf. Clark, 1969a, p. 391), marked and unmarked adjectives do differ in their "breadth of applicability in more comparatives," and the unmarked form is always broader. Second, Huttenlocher and Higgins alleged that "Clark's first principle [the primacy of functional relations] asserts that comparatives with 'good' and 'bad' have different underlying 'meanings,' and no specification is given of any relation between these meanings [p. 501]." But the discussion on lexical marking in Clark (1969a) proposes coding nominal good as [+Evaluative [Polar]], contrastive good as [+Evaluative [+Polar]], and contrastive bad as [+Evaluative [−Polar]]; these are quite explicit specifications of the relations between the meanings of "good" and "bad." This point is expanded in Clark (1970a): "Although sentences (a) [*John is better than Peter*] and (b) [*Peter is worse than John*] have different presuppositions, one can translate from one set of presuppositions to another. Since both *good* and *bad* lie on the superordinate goodness scale, it is possible to say that *worse* means 'less good than' and *better* means 'less bad than' [p. 298]."

To summarize, Huttenlocher and Higgins' criticism of the principle of functional relations is based on misrepresentations of the analysis of comparatives in Clark (1969a) and elsewhere. Furthermore, their rejection of this principle fails empirically, for the data supporting congruence also fully support this principle.

The "Compression" Strategy

It was tentatively suggested in Clark (1969a) that subjects might be using the strategy of "compressing" premises—for example, shortening "A is better than B" implicitly to "A is the better one"—in solving the problems. This suggestion may well be incorrect, yet not for the reasons Huttenlocher and Higgins noted. Each piece of evidence they cited is inconclusive. Consider problems with the premise combination (*a*) "A is better than B, and C is worse than B" and its inverse, and (*b*) "B is worse than A, and B is better than C," and its inverse. As Huttenlocher and Higgins pointed out, compression predicts that *a* will be harder than *b*, but the principle of congruence predicts the reverse. Since *a* was easier than *b* in Clark (1969a) and elsewhere, they rejected compression. But this conclusion does not follow, for compression and congruence could be operating against each other in these premise combinations, with congruence having the larger effect. Huttenlocher and Higgins also argued against compression by pitting two problems favored by compression against two problems disfavored by it. The latencies they cited from Clark (1969a) appeared to show the favored problems to be harder than the disfavored, but the more precise error percentages from Clark (1969b), which they did not cite, show just the reverse, both for comparative problems (26% to 36% errors) and for negative-equative problems (31% to 40% errors). Also, they said nothing about the "indeterminate" problems in Clark (1969b) where it was found that compression was supported in all eight pairs of premise combinations.

Yet there are deviations from compression in the data of DeSoto, London, and Handel (1965) and Hunter (1957). The counterevidence is found in premise combinations like (*a*) "A is better than B, B is better than C" as against (*b*) "B is better than C, A is better than B." Although compression predicts *b* to be easier, DeSoto et al. and Hunter both found *a* to be easier in certain cases. However, this evidence is also in full contradiction to the imagery theory, which also predicts that *b* should be easier. Therefore, the notion of compression, which could hardly be said to "bear the major burden in predicting differences between alternative forms of the three-term series problem [Huttenlocher & Higgins p. 502]," remains, but with less than overwhelming support.

CONCLUDING REMARKS

To sum up, Huttenlocher and Higgins' new theory of reasoning is defective, for they have failed to establish either that the linguistic theory is in need of revision or, more seriously from their point of view, that the imagery theory has any basis in fact. But, as they have correctly pointed out, the linguistic theory is not complete. For one thing, it does not fully specify how information from the two premises are combined. It is laudable that they have tried to fill this gap, particularly by appealing to the visual imagery many people report using in these problems. Yet to repeat Clark (1969a), "the only firm conclusion we can draw at this time is that it has not been demonstrated that the use of spatial imagery differentially affects the solution of three-term series problems [p. 402]." A recent paper[4] proposes quite a different approach to this incompleteness in the linguistic theory. With new evidence, this paper presents an extension of the theory which is compatible both with the present linguistic theory and with the fact that at least some people report imagery in solving three-term series problems.

---

[4] H. H. Clark. More on linguistic processes in deductive reasoning. In preparation.

REFERENCES

CHASE, W. G., & CLARK, H. H. Mental operations in the comparison of sentences and pictures. In L. Gregg (Ed.), *Cognition in learning and memory*. New York: Wiley, in press.

CHOMSKY, N. *Aspects of the theory of syntax*. Cambridge: M.I.T. Press, 1965.

CLARK, H. H. Linguistic processes in deductive reasoning. *Psychological Review*, 1969, **76**, 387–404. (a)

CLARK, H. H. The influence of language in solving three-term series problems. *Journal of Experimental Psychology*, 1969, **82**, 205–215. (b)

CLARK, H. H. Comprehending comparatives. In G. Flores d'Arcais & W. J. M. Levelt (Eds.), *Advances in psycholinguistics*. Amsterdam: North Holland Press, 1970. (a)

CLARK, H. H. How we understand negation. Paper presented at the Committee on Basic Research in Education Workshop on Cognitive Organization and Psychological Processes. Huntington Beach, California, August 19, 1970. (b)

CLARK, H. H. Semantics and comprehension. In T. A. Sebeok (Ed.), *Current trends in linguistics*. Vol. 12 *Linguistics and adjacent arts and sciences*. The Hague: Mouton, in press.

DESOTO, C., LONDON, M., & HANDEL, S. Social reasoning and spatial paralogic. *Journal of Personality and Social Psychology*, 1965, **2**, 513–521.

DOHERTY, P., & SCHWARTZ, A. The syntax of the compared adjective in English. *Language*, 1967, **43**, 903–936.

GREENBERG, J. H. *Language universals*. The Hague: Mouton, 1966.

HUDDLESTON, R. D. More on the English comparative. *Journal of Linguistics*, 1967, **3**, 91–102.

HUNTER, I. M. L. The solving of three-term series problems. *British Journal of Psychology*, 1957, **48**, 286–298.

HUTTENLOCHER, J. Constructing spatial images: A strategy in reasoning. *Psychological Review*, 1968, **75**, 550–560.

HUTTENLOCHER, J., & HIGGINS, E. T. Adjectives, comparatives, and syllogisms. *Psychological Review*, 1971, **78**, 487–504.

HUTTENLOCHER, J., HIGGINS, E. T., MILLIGAN, C., & KAUFFMAN, B. The mystery of the "negative equative" construction. *Journal of Verbal Learning and Verbal Behavior*, 1970, **9**, 334–341.

JONES, S. Visual and verbal processes in problem-solving. *Cognitive Psychology*, 1970, **1**, 201–214.

JUST, M., & CARPENTER, P. Comprehension of negation with quantification. *Journal of Verbal Learning and Verbal Behavior*, 1971, **10**, 244–253.

LEES, R. B. Grammatical analysis of the English comparative constructions. *Word*, 1961, **71**, 171–185.

LYONS, J. *Introduction to theoretical linguistics*. Cambridge: Cambridge University Press, 1968.

SMITH, C. W. A class of complex modifiers. *Language*, 1961, **37**, 342–365.

TRABASSO, T., ROLLINS, H., & SHAUGHNESSY, E. Storage and verification stages in processing concepts. *Cognitive Psychology*, 1971, **2**, 239–289.

(Received August 4, 1971)

Psychological Review
1971, Vol. 78, No. 6, 515–527

JUDGMENT AND RESPONSE IN MAGNITUDE ESTIMATION[1]

JOHN ROSS[2] AND VINCENT DI LOLLO

University of Western Australia

Context is shown to influence not only judgment but, independently, response in the task of magnitude estimation. Pure response effects are isolated and characterized in several studies employing a common shift design and are contrasted with judgmental effects. Response scales in magnitude estimation are presented as interval scales constructed from a direct comparison of stimulus differences and shown then to account for context effects. Implications for psychophysical theory are considered.

In an earlier report (Ross & Di Lollo, 1968a), we presented experimental evidence to show that judgments of heaviness are formed on a complex basis, and that the basis changes with context. One interpretation was that context causes a change in the way heaviness is defined, by changing the emphasis placed on relevant stimulus properties such as weight, density, and size. The general conclusion we drew was that the immediate problem for psychophysics is to try to discover not the form of the psychophysical function for a given attribute, but the laws that determine the basis on which judgments are made. Anderson (1970) reaches a similar conclusion on other grounds.

Our primary concern in what follows is to show that context has powerful effects not only on judgment but also on response scaling in the task of magnitude estimation. The conclusion we reach, partly from new experimental evidence and partly from evidence in previous studies, is that magnitude estimates are made on an interval scale, with a unit and an origin that change with context. What this means is that judgments that themselves are dependent on context will be represented on a response scale which is also susceptible to context effects of a different kind. The double effect of context, first on the judgmental process and second on the response scale, makes the search for a simple psychophysical law for magnitude estimation seem premature until each effect is clearly understood.

ISOLATING RESPONSE SCALE EFFECTS

The distinction between judgmental effects and response scale effects, to which we have already alluded, may be stated as follows: The judgment of stimulus magnitude is the psychological fact that a judge must strive to represent by a number or, generally, a response. Judgments, in our view, depend on the combination of several relevant stimulus properties. The judgmental effect of context is to change the combination of properties on which judgment depends (see Ross & Di Lollo, 1968a). The response scale effect of context is a change or modification in the scale that a judge constructs in order to represent his judgments. Where context affects response scale, but not judgment, the psychological fact remains unchanged, but the way in which it is represented is altered. The subject will use a different number, a different category, or somehow modify the response he would otherwise have made.

We first make a necessary working assumption to enable us to minimize judgmental effects so as to observe response scale effects in a relatively pure form. The assumption is that it is possible to use multidimensional scaling of estimates of difference between stimuli in order to separate classes of stimuli into those for which the effects of context on judgment are weak and those for which they are strong.

Donovan and Ross (1969) have shown that estimates of difference in heaviness

---

[1] This research was supported by Grant ARG A68/16810 from the Australian Research Grants Committee to both authors.

[2] Requests for reprints should be sent to John Ross, Department of Psychology, University of Western Australia, Nedlands, Western Australia 6009, Australia.

produce structures that are not unidimensional, which supports the suggestion (Ross & Di Lollo, 1968a) that lifted weights are judged as complex stimuli and not in terms of the single stimulus property, weight. Other studies in our laboratory have shown that there are, on the other hand, classes of stimuli for which multidimensional scaling of difference estimates does produce unidimensional structures. Where it does so, we take it to mean that a single stimulus property underlies the judgments of difference. Since we argue that the *judgmental* effects of context is to vary the emphasis placed on different stimulus properties, it follows that where only one property is ever considered, context will have no effect.

The conditions under which estimates of difference with respect to a property (like heaviness, length, pattern density, etc.) will produce structures that are not unidimensional are different from those under which estimates of overall dissimilarity for the same stimulus objects will do so. In the former case, the instruction is to ignore everything except the relevant property, whereas, in the latter, all perceptible differences are grist for the judgmental mill. Estimates of differences with respect to a property will produce nonunidimensional structures only if, so to speak, instructions are disobeyed.

Looking at the point more technically, the distinction may be made by considering how difference estimates are formed in each case. Each stimulus object may be considered as a vector of attributes. The instruction to judge the overall difference between two objects x and y is an instruction to estimate a distance value d_{xy}, where $d_{xy} = f(\mathbf{x},\mathbf{y})$, that is, some function of the wo vectors of attribute values. Analysis f a matrix of d_{xy} values for a set of objects ocates them as vectors in a multidimenional space, according to standard multiimensional scaling theory if the requireents of that theory are met.

Estimates of difference with respect to a roperty should be one of two things. The rst possibility is that they are estimates difference considering only one attribute, of the vectors **x** and **y**. That possibility an be represented by saying that the estimates represent distances d^a_{xy} where $d^a_{xy} = f(x_a,y_a)$, x_a and y_a being the values from the vectors **x** and **y** for attribute a. Analysis of a matrix of d^a_{xy} values should produce a unidimensional structure, since the x_a and y_a are scalars.

The second possibility is that the property in question, like heaviness, may depend on two or more stimulus attributes, like weight and density. In this case there should be a two-step estimation of difference with respect to the property, Step 1 being judgment of the property, p, and Step 2 a consequent derivation of a difference estimate. The whole process may be represented as:

$$\begin{array}{ll} \text{Step 1} & p_x = h(\mathbf{x}) \\ & p_y = h(\mathbf{y}) \\ \text{Step 2} & d^p_{xy} = f(p_x,p_y) \end{array}$$

A matrix of d^p_{xy} values should produce a unidimensional structure even where the property p is based on the vectors **x** and **y**, since the values p_x and p_y (or $h(\mathbf{x})$ and $h(\mathbf{y})$) are again scalars.

Failure to find unidimensionality means either that the function h or the function f varies with the pair of objects, x and y, judged. Variation in f may be excluded both as unlikely and tolerable under reasonable assumptions of monotonicity if nonmetric scaling procedures are employed. Finding a structure that is not unidimensional may be taken to mean that the function h varies with the pair of objects judged.

Our assumption is that where several attributes form the basis of judgment of a property, the basis shifts with the pair of objects judged so as to emphasize the attributes on which they differ most and so facilitate discrimination between them. If that is true, h will vary with the pair of objects judged and the matrix of d^p_{xy} values will not produce a unidimensional solution. One specific way in which variation might occur is suggested by Ross and Di Lollo (1968a) in their discussion of the location of judgment lines for restricted ranges of lifted weights. It might be noted that although d^p_{xy} values will produce multidimensional structures if the h function varies, the structure is not a direct recovery of the attribute vectors for the stimulus objects. This fact severely complicates the problem of describing the relationship between the multidimensional structures derived from estimates of difference with respect to a property and scales derived from direct estimates of the property itself.

We rely on multidimensional scaling to determine whether judgments for a class of stimuli depend on more than one property or only on one single property. In the first case, we classify the stimulus class as one where the judgmental effects of context may be observed. In the second case, we classify the stimulus class as one where the judgmental effects of context will be weak or absent. Any response scale effects of context in the second case will be unconfounded with judgmental effects.

Multidimensional scaling, then, enables us to isolate classes of stimuli not at all or only weakly susceptible to judgmental effects, if our working assumption is correct. The assumption could be justified by an argument based on the findings of Donovan and Ross (1969) and our earlier report (Ross & Di Lollo, 1968a), but we prefer to justify it by showing that context effects for the two classes of stimuli defined by multidimensional scaling are clearly distinguishable and easily explicable by a model for response scale effects where judgmental effects should be weak or absent.

AN EXAMPLE OF RESPONSE SCALE EFFECTS

Random checkerboard patterns, like those used by Di Lollo and Kirkham (1969) fall within the class of stimuli that are judged in terms of one property, in this case, pattern density. We report the re-ults of a study using checkerboard pat-erns to show pure response scale effects. n this study, as in all the studies that we onsider, a common experimental design, lled the shift design, was employed.

The shift design employs two comparison imulus series, Series A and Series B, and single standard presented once only be-re judgments of either series begins. he standard is set to a numerical value, rmally 100, and the subject makes magnitude estimates of the stimuli in Series A for a fixed number of trials. In the studies to be reported, each series consists of five separate stimuli, and a trial consists of the presentation of all five stimuli in the series in random order. The order is independently randomized for each trial but is used for all subjects in all groups. On a predetermined trial, Series B is substituted for Series A, abruptly and with no warning to the subject. If Series B is in a higher region than Series A, the shift is upward, if it is in a lower region, the shift is downward. Identical or other series may be used for purposes of control. Suitable controls permit the design to be used to study contrast effects produced by shift. The design yields information about preshift and postshift estimation functions, about the immediate effects of shift, and about the stability of estimates over time.

Figure 1 shows the results of a study in which each of three independent groups of 23 subjects (student volunteers) judged one of three different series of five random checkerboard patterns each with respect to proportion of black area. The stimulus objects have been described by Di Lollo and Kirkham (1969). Briefly, they consisted of 50 × 50 matrices of square-element positions having random structure and varying in percentage of black area from 20% to 80% in steps of 5%. Each matrix was projected on a screen as a square having 4-foot sides. The percentage of black squares in the patterns of each series was as follows: light series—20%, 25%, 30%, 35%, 40%; medium series—40%, 45%, 50%, 55%, 60%; dark series—60%, 65%, 70%, 75%, 80%. The experiment consisted of eight preshift and eight postshift trials. The 50% pattern was used as a common standard stimulus for all groups and was called 100. The standard was presented once only before estimation trials began, and never again. During preshift each group judged a separate series; during postshift all groups judged the medium series. The three groups can thus be designated as light-medium, medium-medium, and dark-medium, respectively.

The data points shown in Figure 1 represent the mean judgment for each pattern

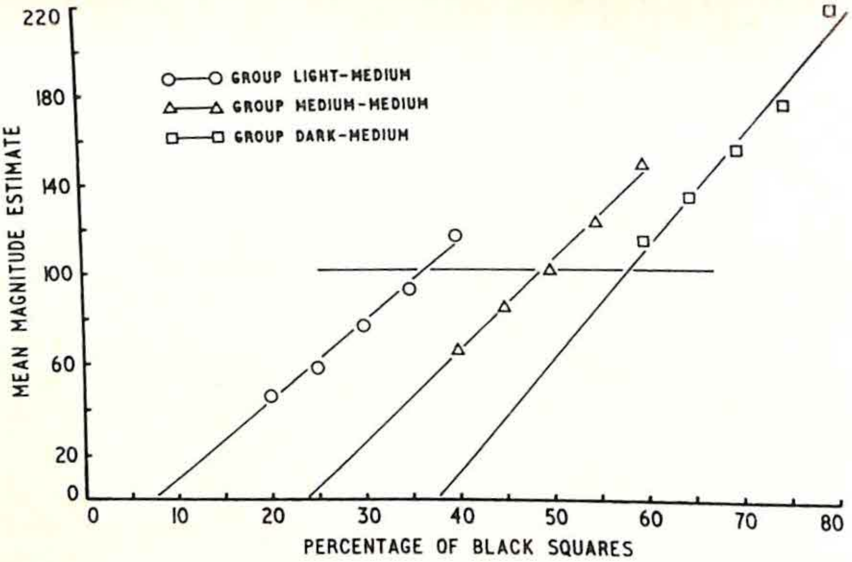


FIG. 1. Mean judgment values for each pattern over the last four preshift trials. (The intersecting horizontal line has been drawn at the level of a mean magnitude estimate of 100.)

averaged over the last four preshift trials for each group.

The groups differ systematically in the stimulus value that they would call zero, as can be seen by producing linear fits to the estimation functions to the abscissa. Second, they differ in the slopes of their estimation functions which increase with mean stimulus level. Third, they differ in the stimulus value that would be estimated as 100, that is in their use of the standard, but the difference is less than for the zero-valued stimuli.

AN EXPLANATION FOR RESPONSE SCALE EFFECTS

The most obvious effect of context is to create an overlap in the range of numbers assigned by subjects to stimuli in adjacent, but not overlapping, stimulus series. The light series group judges a 40% black pattern as 115. The same 40% black pattern is judged as 75 by the medium series group, which would place it, on the light series scale, as a pattern between 25% and 30% black. The two groups are clearly working on different response scales despite the fact that both might be expected to have adopted the same ratio scale, its modulus set by the assignment of the number 100 to a pattern of 50% black squares in both cases.

Similarly, the medium series group judges a 60% black pattern as 146, whereas the dark series group judge it as 114, a value which, on their scale, the medium series group would apply to a pattern about 53% black. The response scale changes as the context moves upward, and the effect is much too large and much too systematic to be dismissed as a mere source of perturbation. As we show later, there are cases where the effect of context is small, but here it is not (cf. Stevens, 1966). It is something to be explained, and we propose an informal model to do so.

The two most important clues in a search for an explanation are provided by the results themselves. As Figure 1 shows, the three groups differ markedly in their description of the standard and in the stimulus that they would describe with the number 100. For the middle group, the medium series group, that stimulus corresponds very closely to the standard. The medium series group does in fact judge the standard, since their stimulus series includes a 50% black pattern. Even after many trials, they judge it as about 100. They can be said to remember the standard

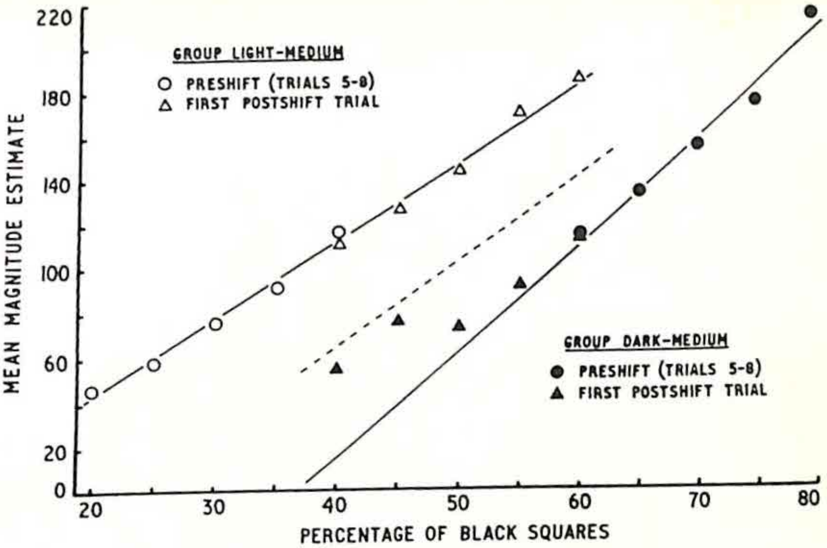


FIG. 2. Mean magnitude estimates for terminal preshift and initial postshift stages for the light-medium and dark-medium groups. (The broken line was fitted through five data points (not shown) representing the mean performance of the medium-medium group over Preshift Trials 5–8.)

accurately. The lower group, the light series group, does not judge the standard, since their stimulus series does not include it, but by extrapolation it can be concluded that the standard would be judged as much higher than 100. When the 50% black standard is encountered on the first postshift trial, it is in fact judged as 144, as can be seen in Figure 2. The number 100 is used to describe a stimulus with a black area much lower than 50%, somewhere about 37%. The light series group can be said to have modified the remembered standard downward, in the direction of the stimulus series within the context of which they make their judgments.

The dark series group is forced in the opposite direction. Subjects would describe the standard as 60, rather than 100, and use 100 to describe a stimulus of some 58% black area. They modify their memory of the standard upward, rather than downward. The first thing the results show, then, is a systematic modification by context of the standard as it is remembered by groups working within a low or a high context, and no modification for the group working within a context to which the standard is central.

The second clue is that the groups differ in the stimulus values which would fall at the zero points on their response scales. The three stimulus values are: light series, 7% black; medium series, 24% black; and dark series 38% black. These values are calculated by extrapolation downward of the linear functions fitted to the data points in Figure 1, and are to that extent suspect. However, any reasonable functions fitted to the three sets of results would show wide differences between the groups in the stimulus values at which the functions meet the abscissa.

Correlations were calculated between log judgments and log stimulus values for individual subjects using data from Trials 5–8. A ratio was computed, for each subject, of the log-log correlation to the correlation for the untransformed values. The average ratios for the light, medium, and dark groups were .982, .953, and .947, respectively, indicating that the raw value regression is marginally more linear than the log-log regression. Linear functions, therefore, fit the data as well as power functions of any other exponent. A further justification for linear fits is the fact that upward linear extrapolation of the light series curve neatly collects estimates by that group for the medium series stimuli on the first postshift trial (Figure 2).

The two clues suggest that context influences both the memory of the standard and the stimulus value that the subject supplies for himself as a lower bound implicit reference stimulus, the stimulus that he would call zero. As context moves up and down the stimulus range, the remembered standard and the lower bound stimulus are forced up or down, with results that can be observed in Figure 2.

We believe that the subject in the task of magnitude estimation constructs his estimates for stimuli indirectly, from estimates of stimulus difference. Subjects, when asked, say that they compare each stimulus with others, especially the last one or two, and adjust their previous estimates upward or downward by an appropriate amount. We are inclined to believe them, and to take them to mean that they estimate a numerical value for a stimulus difference, then construct their estimate for the stimulus from the stimulus difference value and previous estimates. What subjects describe is interval scale behavior, that is, the direct representation of intervals as numbers by a process of comparison of intervals, and the construction of numbers to represent stimuli from the numbers for stimulus differences.

The informal model we propose assumes that subjects are sensitive to stimulus differences, and estimate directly, without the use of numbers, the ratio of one difference to another. They can, so to speak, estimate how many times a larger difference would contain a smaller. If we grant that capability, and if we also grant a memory adequate to enable the comparison of a present with a previous stimulus, and of two previous stimuli, then we have enough to explain how interval scale representations can be constructed.

The subject starts with a standard stimulus, s, and the instruction to represent it by a numerical value, n. Suppose that he imagines a lower bound stimulus, z, the stimulus he would call zero. If he can sense the difference between s and z, d_{sz}, then he can also use n to represent the difference d_{sz}. We now present him with a second stimulus, r, and he senses the difference value between r and s, d_{rs}. Granting him the ability to estimate the ratio of stimulus differences directly, he can calculate a numerical value for the difference d_{rs} by applying the ratio of d_{rs} to d_{sz}, which is obtained directly, to the number, n, that represents d_{sz}. Let us say that the result is the value a. The adjustment of n, the value that represents both the difference d_{sz} and the stimulus value s, upward or downward by a, yields a value m that represents the new stimulus, r. A direct comparison of differences thus allows the direct representation of a difference, and the indirect representation of a stimulus. The subject is fortunate enough to have a check. He can compare d_{rz} with d_{sz}, and repeat the process to see if he arrives at the same result.

As new stimuli are introduced, the process can proceed in the same fashion, but with more options and with more opportunities for checks. A fresh stimulus can be compared with the previous one or two, with the standard, and with the lower bound; whatever is done, the result should be the same.

What context does, we believe, is introduce a systematic bias into the stimulus value the subject remembers as the standard, s, and the stimulus value he imagines for the lower bound stimulus, z. The effect is to produce interval scales with different intercepts and different slopes.

To illustrate, it is sufficient to consider the medium and dark series groups judging the stimulus they have in common, the 60% black square. The medium series group remembers the standard veridically, but imagines the lower bound stimulus as about 24% black. For them, the number 100 means both a 50% black pattern and the difference between a 50% and a 24% black pattern. They perceive the difference between the 60% black stimulus and the standard as being contained a little more than twice within the difference between the standard, s, and the lower bound, z. In the notation used above, d_{sz} is a little more than twice as big as $d_{60,s}$. Since d_{sz} is represented by 100, the appropriate number for $d_{60,s}$ is a little under 50, that is, the number that bears the same ratio to 100 as $d_{60,s}$ does to d_{sz}. The ratio is the result of a direct comparison of differences; the appropriate number for the new difference is the result of a numerical calculation. The stimulus pattern of 60% black is judged as 100, the number that represents the standard, plus a little less than 50, the numerical value that represents the difference $d_{60,s}$. It is judged by the medium series group as 146.

The dark series group remembers the standard as a pattern of about 58% black squares, and places the lower bound at about 38% black. The 60% black square is only a little different from the standard as remembered. The difference, $d_{60,s}$, is represented by a number just above 10, and the 60% black square is called 114.

It is not our intention to develop the sketch of the representation process into anything so ambitious as a formal measurement theory, but an attempt to do so might proceed along the lines of a psychophysical relation theory constructed by Krantz (1970). Krantz assumes that the fundamental object judged is "the perceived relation between a pair of stimuli," and that ratio between stimuli can be apprehended by a direct mental matching operation, like the physical operation carried out in comparing lengths. He goes on to construct a set of axioms from which it can be proved that representation of stimuli by numbers is guaranteed, and that the permissible representations are related by power transformations. What we have assumed about stimulus differences is very close to what Krantz assumes about stimuli themselves, though we have probably assumed sufficiently more to require multiplicative rather than power transformations.

Krantz (1970) proposes a mechanism by which a sensation "ratio," a relationship between members of a pair, is arrived at directly, with no numerical calculations. His assumption is that pairs of properties, like heaviness or loudness, are matched with length pairs, and then one member of the pair "fitted into" the other as many times as it will go. The process of mentally "fitting in," like the physical processing of laying one length off against another, yields a ratio directly. Numerical calculation enters in only when a person assigns a number n to a stimulus y_i. The number he assigns to y_i is $n = pr$ if the number he has assigned to x_i is p and if the directly obtained ratio of y_i to x_i is r.

Our assumption is that it is pairs of differences, not pairs of properties, that are compared, and one difference that is "fitted into" another, not one object or its length equivalent. If it is reasonable to assume a mechanism that matches object pairs with length pair equivalents, it is as reasonable, and perhaps more so, to accept a matching of difference pairs and length pairs. Length itself may be a direct perception of difference in location between points assigned positions within a spatial system, and rather more naturally coordinated to other perceived differences. The construction of numbers for stimuli from numbers that represent differences entails a further step, the addition or subtraction of a number for a difference to or from a number previously assigned to another stimulus.

RESPONSE SCALES AND THE EFFECTS OF SHIFT

An overlap in the range of numbers assigned by independent groups to adjacent stimulus series can be taken as a primary sign of response scale effects in cases where stimuli fall within the unidimensional class. Overlap is ambiguous for multidimensional stimuli, like lifted weights, since it could be caused by differences in the basis of judgment for stimuli within different contexts.

When independent groups working in different stimulus ranges are shifted abruptly to a common postshift series, their initial estimates should depend on the response scale established in the preshift series. With repeated trials, as the postshift series establishes new reference points, the standard as remembered, and the lower bound stimulus, all groups should tend to establish similar response scales, and the initial effects of shift should begin to disappear.

Returning to the study described in the foregoing, Figure 2 shows the initial effects of shift when groups judging random checkerboards are shifted abruptly to the medium series. Straight lines have been

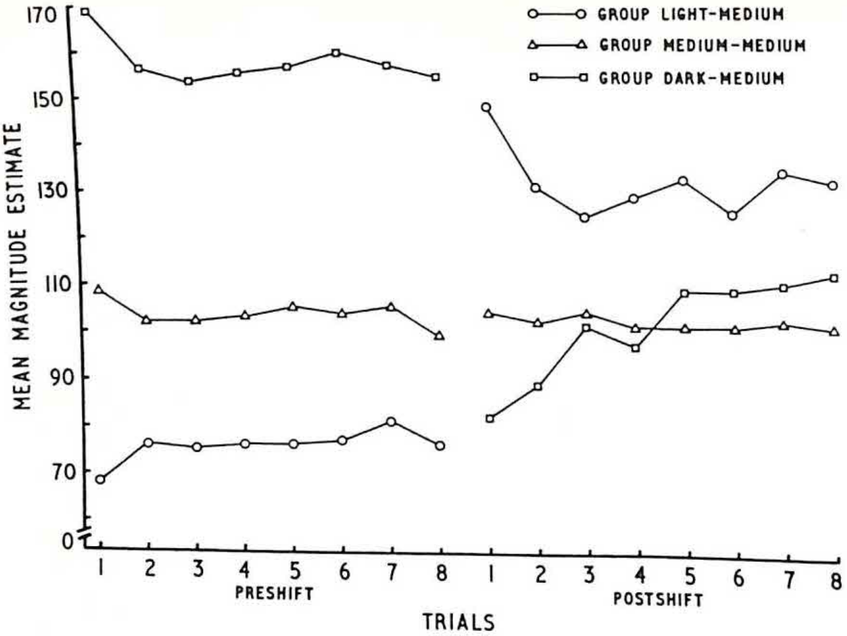


FIG. 3. Mean judgment value for each group on each trial.

fitted through the preshift data of both shifted groups and have been extended over the range of the medium series. On the first postshift trial, the upward shifted group follows the preshift estimation function almost exactly in judging the postshift series. The downward group does not and could not unless some stimuli were assigned unrealistically small numbers. The downward shifted group uses a new scale defined at the upper end by the lowest preshift stimulus, and at the lower end by a new lower bound stimulus, forced abruptly upon it by the change in context. The lower bound is roughly equal to that of the medium series group on its last four preshift trials.

With repeated postshift trials, both shifted groups moved toward the control function, as the postshift series established a new context.

The five responses made by each subject on each trial were averaged to yield the mean response by each subject for the series on each trial. Mean magnitude estimates for each group on each trial, before and after shift, are shown in Figure 3. The point that emerges from Figure 3 is the striking similarity in preshift group differences, in the initial effects of shift, and in convergence toward control levels, to results typically found in similar studies using category estimation (Di Lollo, 1964; Ross & Di Lollo, 1968b). There is also an important difference: The effects of upward and downward shift are fairly symmetrical when category estimation is used. Here they are not. Upward shift produces a larger initial deflection from the control level than downward shift. Further, the upward shifted group moves toward but does not reach the control level with repeated postshift trials, whereas the downward shifted group not only reaches but moves past the control level. We show later that the asymmetry of upward and downward shift effects is not an invariable feature of magnitude estimation but depends on the range of numbers available as permissible responses.

So far it has been shown that where overlap between preshift estimates is large, shift effects are also large and consistent with our explanation for response scale effects. Results for two further cases, length of lines and pitch of tones, are now reported. Both fall within the unidimensional class, though there is some evidence

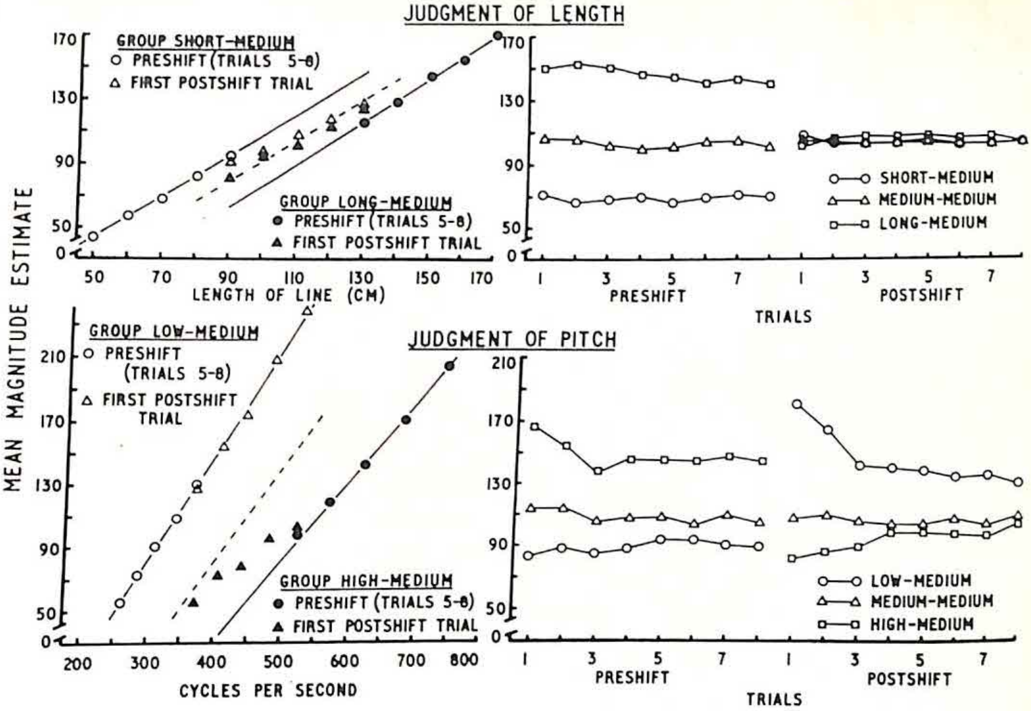


FIG. 4. Mean magnitude estimates for judgment of length of line and pitch of tone.

for minor judgmental complications in the case of tones.

The procedural and methodological details of the two experiments were as follows:

Length of line. Three independent groups of 20 subjects each (student volunteers) made magnitude estimations of the length of one of three different series of five straight lines each. The lines were projected as horizontal strips of light having 2 centimeters breadth and variable length about the center of a flat screen 7 meters wide and 2 meters high, located in a semidark room. The length of each line in each series was as follows (in centimeters): short series—50, 60, 70, 80, 90; medium series—90, 100, 110, 120, 130; long series—130, 140, 150, 160, 170. The subjects viewed the lines, presented one at a time, from a distance of approximately 8 meters. The 110-centimeter line was used as a common standard for all groups and was called 100. All other procedural details were the same as for the experiment with random checkerboards, including a shift of the medium series after eight preshift trials.

Pitch of tone. Three independent groups of 40 subjects each (student volunteers) made magnitude estimations of the pitch of one of three different series of five tones each. The tones had been recorded on magnetic tape and were presented through headphones at an intensity of 72-decibels sound pressure level. The frequency of each tone in each series was as follows (in cycles per second): low series—262, 285, 311, 339, 370; medium series—370, 404, 440, 480, 523; high series—523, 571, 622, 679, 740. Since the equal-loudness contour does not vary more than 3 decibels over the range of frequencies at the sound pressure level used in this experiment, no correction for loudness was made. The 440-cycles-per-second tone was used as a common standard for all groups and was called 100. The subjects were instructed that a tone having pitch twice as high as the standard should be called 200, a tone having pitch half as high as the standard should be called 50, and so on. All other procedural details were the same as for the experiment on length of line, including a shift to the medium series after eight preshift trials.

Figure 4 shows the results of both studies. The overlap in the preshift phase between adjacent stimulus series is small in the case of judgment of length, and large in the case of judgment of pitch. Points for individual stimuli in the preshift phase are shown in the left panels of Figure 4 only for the top and bottom series, but overlap between adjacent series can be estimated from the straight lines fitted to both sets of top and bottom series and drawn without individual points being shown for the middle series.

The left panels of Figure 4 also show the performance of both shifted groups on the first postshift trial. In judgment of length, both the upward-shifted group (short-medium) and the downward-shifted group (long-medium) move to the control level immediately. Both sets of points fall almost exactly on the control curve. In judgment of pitch, the upward-shifted group (low-medium) follows the preshift curve on the first postshift trial, as also happened in judgment of black area with random checkerboard patterns. The downward-shifted group (high-medium) also behaves like their counterparts judging checkerboard patterns. The top end of the curve is tied to the preshift scale, but there is an abrupt readjustment in slope to accommodate the lower stimuli.

The right panels of Figure 4 show the effect of continued exposure to the postshift context. Judgment of length remains stable over the whole postshift phase, showing neither an initial deflection from control levels in response to shift, nor any drift in level with repeated postshift trials. Judgment of pitch shows a pronounced deflection upward with upward shift and downward with downward shift. With repeated trials, both shifted groups drift toward the control level. The asymmetry between upward and downward shift that was observed with checkerboard patterns can also be seen clearly with pitch.

The results of the two experiments show that where overlap between numbers assigned to adjacent stimulus series is small, the effects of shift are small, and that where overlap is large, the effects of shift are large. This correlation of effects is to be expected if both preshift and postshift effects depend, as we say they do, on sensitivity of reference stimuli to context.

What it is that determines the sensitivity of reference stimuli to context is another question. We do not offer any firm opinion, but we venture to suggest that the primary determinant is whether or not people have had training in the mensuration techniques for the stimulus property concerned. Everyone has been taught the principles of mensuration for length and weight. The concepts of zero length and zero weight are well established, the operations to increase or reduce length and weight are known, and elaborate reference systems for both length and weight are part of the intellectual equipment of subjects like those who served in our experiments. Pitch is another matter. The concept of zero pitch is not a clear one, the operations to increase or reduce pitch are not familiar, and there is no well-established reference system for pure tones as there are for stimulus domains such as length and weight. While the *concept* of zero pattern density may be clear enough, the operations to increase or reduce the pattern density of a random checkerboard are, once again, obscure to most people, and there is no well-established and highly overlearned stock of exemplars to remind a judge of the look of a stimulus having a given absolute density under the conditions of the experiment.

Our suggestion is that thorough familiarity with concepts, operations, and exemplars, all of which are part of training in mensuration, minimize context effects by providing a general context within which the experimental context falls into place. Where familiarity is lacking, the experimental situation is prone to distort the general context, producing the context effects that are observed.

RESTRICTION OF RANGE

There is an asymmetry in the examples considered so far between the effects of shift on the first postshift trial following upward and downward shift. Estimates of upward-shifted groups follow the preshift estimation function, whereas estimates of downward-shifted groups do not. The

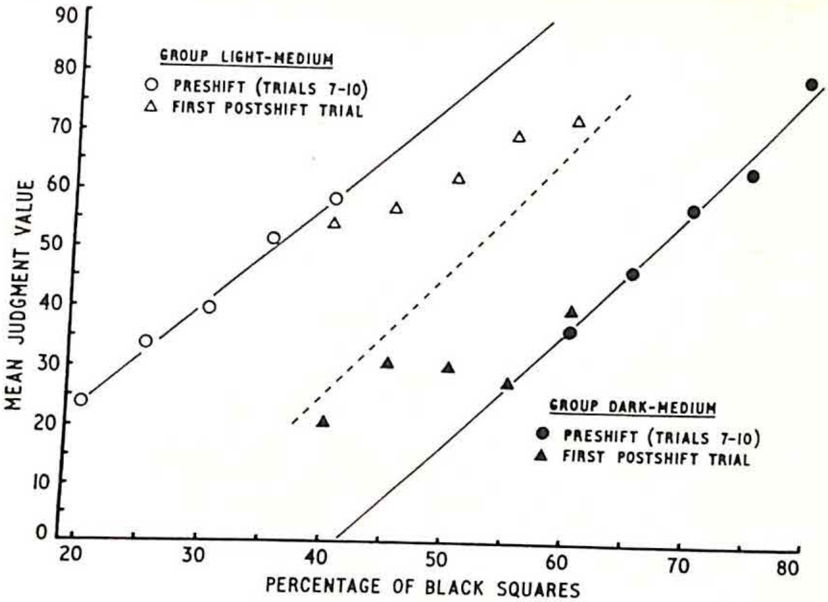


FIG. 5. Mean judgment values for terminal preshift and initial postshift stages for the light-medium and dark-medium groups. (The broken line was fitted through five data points (not shown) representing the mean performance of the medium-medium group over Preshift Trials 7–10.)

explanation given is that if the preshift function were to be maintained following downward shift, some stimuli would be assigned unrealistically low numbers and judges are therefore forced to make an abrupt readjustment. If the explanation is correct, symmetry between upward and downward effects should be found when there is a maximum as well as a minimum valued stimulus within a domain, and where both have a fixed value, like 0 and 100. Di Lollo and Kirkham (1969) report such a case in a study of pattern density where all black and all white checkerboards were maximal and minimal stimuli, respectively; one defined as 100, the other as 0. Only a portion of the experiment need be considered here. Three groups (light-medium, medium-medium, dark-medium) of 13 subjects each were required to estimate the percentage of black area in each of five patterns under conditions that were otherwise identical to those of the study with checkerboard patterns described in the foregoing. Further analysis of the data reported by Di Lollo and Kirkham (1969) reveals symmetrical shift effects on the first postshift trial as shown in Figure 5.

Downward-shift effects are exactly like those reported for the previous checkerboard study (Figure 2) and for the experiment with tones (Figure 4, lower left panel). Upward-shift effects are different from those reported above and symmetrical with downward-shift effects in that the estimate for the lower stimulus of the postshift series is fixed by the preshift estimation function, but the first postshift estimation function has a much shallower slope, as it must have if responding with unrealistically high numbers is to be avoided.

The estimation of black proportion for checkerboard patterns is a task where the stimuli are unidimensional and where familiarity with mensuration techniques is low. People have typically not been taught to estimate black proportion as they have been to estimate length and weight. Where subjects are free to use any positive number to represent their judgments, there is an asymmetry between the effects of upward and downward shift, as was shown earlier. The effects become symmetrical when the task is changed to introduce an upper limit as well as a lower limit on the numerical

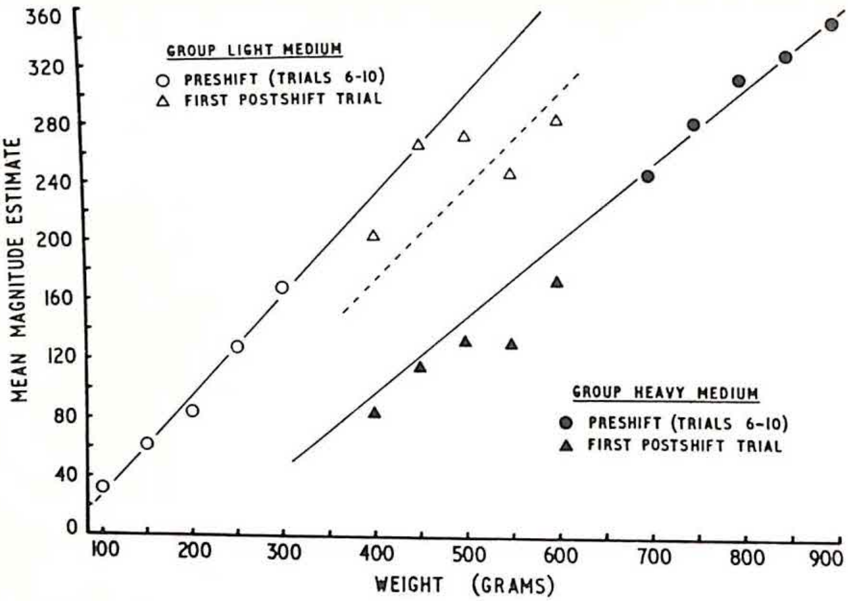


FIG. 6. Mean magnitude estimates for terminal preshift and initial postshift stages for the light-medium and heavy-medium groups. (The broken line was fitted through five data points (not shown) representing the mean performance of the medium-medium groups over Preshift Trials 6–10.)

responses available for the representation of judgments. That a task modification which affects only the response range available to the subject alters the effects of shift so decisively strongly suggests that response effects are the predominant ones in the experimental situation.

JUDGMENT AND RESPONSE EFFECTS

We now show that the effects described above can be clearly distinguished from effects found with lifted weights, stimuli that do not fall in the unidimensional class (Donovan & Ross, 1969). The results shown in Figure 6 represent further analyses of data reported in a previous paper (Ross & Di Lollo, 1968a, Experiment I) to which the reader is referred for procedural and methodological details. Briefly three independent groups of 20 subjects each (student volunteers) made magnitude estimations of the heaviness of one of three different series of five weights each. The weights consisted of 4-ounce polythene containers, 5.3 centimeters in diameter and 7.3 centimeters high, filled with lead shot and paraffin wax to the weight required. The weight of each stimulus object in each series was as follows (in grams): light series—100, 150, 200, 250, 300; medium series—400, 450, 500, 550, 600; heavy series—700, 750, 800, 850, 900. The 500-gram weight was used as a common standard for all groups and was called 100. There were 10 preshift and 10 postshift trials. All other procedural details were the same as for the experiment with random checkerboards, reported in the foregoing, including a shift to the medium series. Although the zero-valued stimulus on the three preshift estimation functions shown in Figure 6 increases systematically with series level, the slopes decrease rather than increase correspondingly. Moreover, the differences in the 100-valued stimuli are greater than, not less than, the differences in the 0-valued stimuli.

Postshift effects are still more discrepant. In the first place, the estimations on the first postshift trial do not follow the preshift estimation function either in upward or downward shift (Figure 6); and, in the second place, the mean estimate for the series continues to rise over postshift trials whether the shift was upward or downward (Ross & Di Lollo, 1968a, Figure 5).

Ross and Di Lollo (1968a) showed that both preshift context effects and postshift effects, including the almost invariable upward drift in level over preshift and postshift trials, could be explained by a model for the complex determination of judgments of heaviness, ignoring response-scale effects. While it is almost certainly true that response effects are also present in estimates of heaviness, it does seem that it is judgmental effects that are predominant, and it does seem that they can be distinguished from response effects.

The classification of stimulus domains into those susceptible to judgmental effects and those not susceptible or only weakly susceptible was based on a working assumption that multidimensional scaling of judgments of stimulus differences would reveal more than one dimension in the former case and only one dimension in the latter. The validity of the classification scheme is substantiated first by the sharp differences between preshift and postshift effects found for the two classes, for example, between lifted weights and checkerboard patterns; second by the fact that a variety of effects can be shown to be correlated within the class of stimuli not susceptible to judgmental effects. Where the overlap in the range of numbers assigned by independent groups to adjacent stimulus series is large, the effects of shift, immediately and with continued postshift trials, are also large. Where the overlap is small, as it is with straight lines, shift effects are also small. Finally, the validity of the classification and the effectiveness of the working assumption are attested by the fact that the effects found within both classes of stimuli can be explained by appropriate models.

REFERENCES

ANDERSON, N. H. Functional measurement and psychophysical judgment. *Psychological Review,* 1970, **77**, 153–170.

DI LOLLO, V. Contrast effects in the judgment of lifted weights. *Journal of Experimental Psychology* 1964, **68**, 383–387.

DI LOLLO, V., & KIRKHAM, R. Judgmental contrast effects in relation to range of stimulus values. *Journal of Experimental Psychology,* 1969, **81**, 421–427.

DONOVAN, R. J., & ROSS, J. The dimensionality of judgments of heaviness. *Perception and Psychophysics*, 1969, **6**, 247–249.

KRANTZ, D. H. *A theory of magnitude estimation and cross-modality matching.* (Tech. Rep. MMPP 70-6) Ann Arbor: Michigan Mathematical Psychology Program, July 1970.

ROSS, J., & DI LOLLO, V. A vector model for psychophysical judgment. *Journal of Experimental Psychology* 1968, **77**(3, Pt. 2). (a)

ROSS, J., & DI LOLLO, V. Category scales and contrast effects with lifted weights. *Journal of Experimental Psychology*, 1968, **78**, 547–550. (b)

STEVENS, S. S. On the operation known as judgment. *American Scientist*, 1966, **54**. 385–401.

(Received March 2, 1971)

Psychological Review
1971, Vol. 78, No. 6, 528–539

METACONTRAST AND LATERAL INHIBITION[1]

BRUCE BRIDGEMAN[2]

Stanford University

A Hartline-Ratliff lateral inhibitory network, modified for application to mammals, is stimulated with patterns having all of the key parameters of stimuli used in metacontrast experiments. Network activity elicited during the application of stimuli follows a monotonic masking function, but a U-shaped function results when activity both during and after stimulation is examined. The longer processing time needed for obtaining U-shaped functions reflects the higher response criterion necessary for obtaining this function psychophysically. Forward masking, simultaneous presentation, and backward masking conditions are illustrated. In addition, the model qualitatively explains mutual masking, decreasing effectiveness of masking with increasing interstimulus separation, and change to monotonic functions with stronger masking stimuli, and it makes some predictions. Metacontrast is interpreted as a handicap which the organism accepts in order to gain the advantages of stimulus processing by lateral inhibition. Psychophysical and physiological evidence is cited to support the model.

Backward masking, the obscuring of a brief visual stimulus by a subsequent visual stimulus, has long been of interest to psychologists because it apparently works backward in time. Backward masking is called metacontrast when a patterned masking stimulus obscures a target pattern even though the two have no spatial overlap. The extensive literature on metacontrast has been recently and ably reviewed by Kahneman (1968), Raab (1963), and Weisstein (1968), while Alpern (1952) has reviewed the earlier literature. Some generalizations from this work can guide attempts to account for metacontrast effects.

The two types of masking function commonly found are the monotonic, in which masking is greatest when the two stimuli are presented simultaneously, and the U-shaped, in which the most effective masking occurs when the masking stimulus begins about 50–80 milliseconds after onset of the target. The latter condition, in which both target and mask are seen if presented simultaneously but the target is obscured if the mask follows it, is the most paradoxical condition and has been the most difficult to explain theoretically.

Many of the conditions necessary for obtaining U-shaped metacontrast functions have been established, though there is disagreement on some points. The function is obtained only when the masking stimulus surrounds the target stimulus but shares no common borders with it; an annulus surrounding a disc or two parallel bars flanking a target bar are common stimulus situations. Metacontrast is a weak effect, usually measured as an increase in the number of errors made in discriminating some characteristic of the target, performance remaining above chance even at the peak of masking. The U-shaped function is best obtained with equal-energy stimuli, and it is observed when a subject must do more than simply detect the presence or absence of the target; he has a higher criterion (Kahneman, 1968). Unlike other forms of visual masking, metacontrast can be obtained dichoptically (Schiller & Smith, 1968), so that the effect cannot be due entirely to retinal or lateral geniculate interactions. Yet it is still crucially dependent on spatial parameters because it disappears when the separation between target and mask increases much beyond 2° of visual angle, and it is best obtained in the peripheral retina. Masking is mutual.

---

[1] Supported by National Institute of Mental Health Grant MH 12970 to Karl H. Pribram. I am indebted to Karl Pribram, L. Ganz, R. W. Phelps, and especially D. N. Spinelli for many helpful suggestions.

[2] Requests for reprints should be sent to Bruce Bridgeman, who is now at the Physiological Institute, Free University of Berlin, Arnimallee 22, 1 Berlin 33, West Germany.

that is, the perception of both target and mask is interfered with to some degree, the target stimulus being more strongly inhibited. A satisfactory theory of metacontrast must account for all of these characteristics.

THEORIES

Weisstein (1968) has divided metacontrast theories into three general groups: confusion, overtake, and inhibition. Confusion theories argue that

> the more similar the mask is to the target stimulus, or the more noise generated by the interaction of the two stimuli, or the more difficult the spatial or temporal resolution, the harder it will be to discriminate the target from the overall stimulus situation [Weisstein, 1968, p. 496].

This hypothesis only restates the problem in more obscure terms and has difficulty in accounting for both monotonic and **U**-shaped functions at the same time. One explanation of the **U**-shaped function (Kahneman, 1968) notes that the geometry of the metacontrast design implies "impossible" apparent motion and the target is therefore suppressed, but there is no reason why the mask should not be suppressed instead. Further, apparent motion can be seen over a larger visual angle than metacontrast suppression.

The overtake hypothesis, stating that the target stimulus is held in a short-term buffer store but is replaced by the masking stimulus before readout, encounters similar problems as well as difficulties in explaining forward masking. Both of these theories generally ignore the uniquely spatial nature of metacontrast, and have difficulty in predicting both monotonic and **U**-shaped functions with the same model parameters.

A third set of theories is neurologically based and uses lateral inhibition to account for the reduced responses to the target stimulus. This kind of inhibition operates in a group of neurons with parallel inputs and outputs (Figure 1), in which each neuron sends out collateral axons to inhibit its neighbors *in the same layer*, either directly or through inhibitory interneurons.

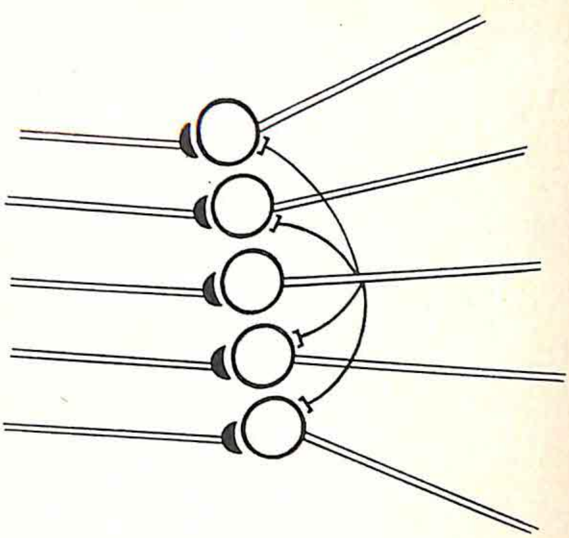


FIG. 1. Diagrammatic representation of lateral inhibition in a neuronal network. (Information flows from left to right. Only a few inhibitory collaterals for a single cell are shown. Inhibition is applied recurrently, so that each cell in the network can be influenced by the firing of its nearby neighbors.)

Inhibitory networks are common in sensory systems, providing an anatomical substrate which is lacking in other explanations of metacontrast. Lateral inhibition has a role in many other visual phenomena, such as light and dark adaptation (Ratliff, Knight, & Milkman, 1970), contour enhancement (Ratliff, Hartline, & Miller, 1963), figural aftereffects (Ganz, 1966), successive contrast (Singer & Creutzfeldt, 1970), and Mach bands (Mach, in Ratliff, 1965). Thus an interpretation of metacontrast in terms of lateral inhibition is parsimonious, invoking only classical neuronal mechanisms. Furthermore, there is physiological evidence for masking effects in lateral inhibitory networks. Ratliff et al. (1963) illuminated a group of *Limulus* ommatidia and showed that the response of a neighboring ommatidium to light was decreased after a time delay. Grüsser, Petersen, and Sasowski (1965) have found metacontrastlike effects due to lateral inhibition in the cat retina, but the effects were too brief (5–40 milliseconds) to be those responsible for metacontrast in humans. Schiller (1968) obtained a monotonic masking function from single cells of the cat lateral geniculate. While these experiments cannot account for human metacontrast because of its dichoptic nature, they do show that lateral inhibition can cause some kinds of backward masking under physiological conditions.

The most complete simulation of metacontrast in the literature is Weisstein's (1968) model, which actually uses a form of forward inhibition with a lateral component. Her two-receptor model cannot simulate the nonadjacency of stimuli required for true U-shaped metacontrast functions (two adjacent stimuli will appear as a single larger pattern if their onset and offset are simultaneous). Though the receptors in her model are identical, the second-layer cells are not, so that the simulation will not work if target and mask are interchanged.

THE MODEL

A stimulus impinging on a network incorporating lateral inhibition causes a series of disturbances to spread away from the stimulus. Activity in both the stimulated locus and in nearby areas of the net becomes modified by the spread of inhibition, and even after stimulation ceases, these oscillations continue to influence the net until they damp out. Such effects have been studied in detail by Hartline, by Ratliff, and by others in the *Limulus* eye (Dodge, Shapley, & Knight, 1970; Ratliff, 1965). On the basis of his observations, Ratliff has developed a mathematical model of the system in which each neuron inhibits neighboring neurons by an amount proportional to its own activity. If r_p is the ongoing firing rate of neuron p, and r_j is the rate of a nearby neuron j, the effect of r_j on r_p is described as

$$r_p = e_p - k_{p,j}(r_j - r_{p,j}{}^\circ) \qquad [1]$$

where e_p is the excitatory input to neuron p, $k_{p,j}$ is an inhibitory coefficient with a value $0 \leq k \leq 1$ expressing the effect of j on p, and $r_{p,j}{}^\circ$ is the corresponding threshold of inhibition. Interactions also travel in the opposite direction, of course, with p inhibiting j. Similar effects connect each neuron with many others, so that the steady-state firing rate r_p is more completely described by

$$r_p = e_p - \sum_{j=1}^{n} k_{p,j}(r_j - r_{p,j}{}^\circ) \quad j \neq p \qquad [2]$$

for interactions with n neurons in a one-dimensional net.

A similar equation can be written for each neuron in the network, yielding a matrix of linear equations describing the entire network's interactions. But the inhibition described in each equation is not truly recurrent, that is, the action of j on p is found without taking into account the action of p on j. In the model, this nonrecurrence represents the passage of time; at any time t a neuron is being inhibited by signals which originated from its immediate neighbors at $t-1$, from twice as far away at $t-2$, etc., so that the basic equation becomes

$$r_p(t) = e_p(t) - \sum_{j=1}^{n} k_{p,j}[r_j(t - |p-j|) - r_{p,j}{}^\circ] \quad j \neq p. \qquad [3]$$

Applying this equation to each neuron in succession yields a new set of activities in the network, to which the same equation can be applied again. After an infinite number of such iterations, the state of the system converges to that obtained in a recurrent one (Figure 2). A further discussion of this and other properties of the model is given by Ratliff (1965, pp. 130–142).

Equation 3 is a special case of the *Limulus* model with a simple time lag of inhibition as presented by Ratliff, Hartline, and Lange (1966), which according to those authors gives qualitatively correct predictions of *Limulus* eye responses in most cases. The present model restricts inhibitory effects to a constant apparent conduction rate, which may be an oversimplification in the *Limulus*, where inhibitory latency is related partly to (*a*) the time needed to reach the inhibitory threshold (Lange, Hartline, & Ratliff, 1966; Ratliff et al., 1966) and partly to (*b*) the activity of the ommatidia exerting the inhibition

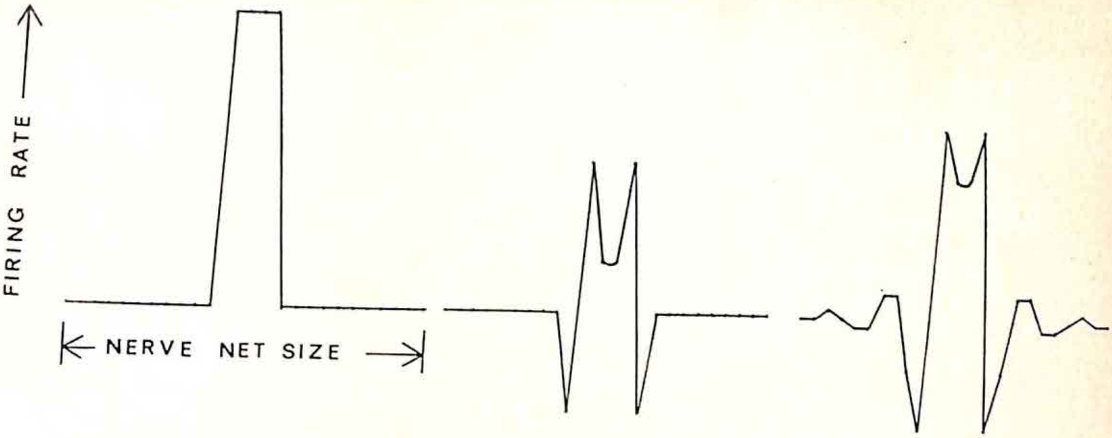


FIG. 2. The effect of repeated iterations of lateral inhibition on a continuously applied stimulus. (The stimulus is shown on the left; after one iteration of inhibition, the network's activity is shown at center. After a large number of iterations, activity converges to that shown at right. The same result is obtained by solving simultaneous linear equations for each neuron in the network.)

(Ratliff et al., 1966). In this application of the model to a mammalian system, inhibitory thresholds are set to zero, and the average activity of cells exerting inhibition is similar in all simulated conditions, minimizing the influences of threshold and activity on apparent conduction time under the simulated conditions, so a constant conduction speed is assumed. Wuttke and Grüsser (1968), in fact, find a constant conduction rate of lateral inhibition in cat retina.

Because this simulation involves assumptions about central structures in the nervous system, no thresholds for lateral inhibition are used. A background level of firing is maintained so that excitation and inhibition can modify the final firing level by means of a dendritic microstructure of *graded* potentials which can register even very small changes (Pribram, 1971, Ch. 1).

Inhibitory coefficients $k_{p,j}$ were chosen to follow the results of Barlow (1969), who found a slight strengthening of inhibition at short interommatidial distances in *Limulus* followed by a decline in effects for more distant elements. The coefficients used here closely follow the shape of Barlow's "vertical" curve (extrapolated to 0), condensing its lateral spread; the first two $k_{p,j}$'s of .3 bracket the peak of inhibition, while further *k*s of .1 and 0.0 follow the decline. The coefficients are larger than those found in *Limulus* (Ratliff et al., 1966) to compensate for the lack of effects from the second dimension in a one-dimensional network. Inhibition is symmetrical on each side of a neuron.

Lange et al. (1966) present a more developed version of the same additive model of spatial and temporal interactions in *Limulus*, where the state of the network is defined for each millisecond of time so that influences can be traced impulse by impulse, and durations of postsynaptic potentials, etc., can be simulated precisely. Some of their assumptions, specifically the generation of spikes and their direct addition to a lateral inhibitory pool, do not apply to a vertebrate system in which lateral inhibition is mediated by interneurons which themselves can be stimulated to change the characteristics of inhibition (Hull, 1968; Kalil & Chase, 1970). Later refinements of the additive system (Knight, Toyoda, & Dodge, 1970; Ratliff, Knight, & Graham, 1969; Ratliff et al., 1970) have modeled temporal effects more generally, in formats appropriate for analyzing responses to sinusoidal stimulation.

APPLICATION TO METACONTRAST

The following simulations use a revision of a PDP-8 computer program written by R. W. Phelps and D. N. Spinelli (unpublished). Using Equation 3 to calculate lateral inhibitory effects, it prints the network's state after each iteration. Stimuli can be added or removed after each iteration. To avoid instabilities, the ends of the network are tied together to form a circle, and simulations were stopped when effects from stimuli at the center spread to the ends, preventing effects from intersecting across the "seam." The program solves Equation 3 for each neuron in the network successively; the response $r_p(t)$ is printed out for each neuron and stored. For the next time iteration, the $r_p(t)$ values are moved to $r_p(t-1)$ and used to evaluate $r_p(t-1)$ for effects on adjacent cells; in like manner, earlier matrices $r_p(t-n)$ are stored for use in calculating interactions with more distant cells.

With this program, the effects of lateral inhibition on a brief stimulus can be observed directly by a three-stage process. First, the network is "adapted" to a steady state by giving each of the neurons a "background" stimulation level e_p for several iterations. (It was discovered early that one must simulate all stimulus conditions, not just the interesting ones!) After adaptation, a stimulus is applied by increasing the input values for some neurons for a few iterations, and finally the stimulus is "turned off" by reassigning the background stimulation level to all neurons, the network again going through successive iterations of inhibition.

Stimulus parameters for the metacontrast simulation were chosen to follow as closely as possible those used in behavioral masking experiments. If the one-dimensional network is considered as a cross section through the center of a stimulus array, a disc would appear as a step increase in the excitation of some elements (Figure 2, left). In the 24-cell array used here, the center four cells represent the target stimulus of metacontrast studies. The masking stimulus, a cross section of an annulus, consists of two groups of two neurons each separated by one neuron from an edge of the disc. The target and mask have equal area, intensity, and contrast, so that all of the parameters of metacontrast stimuli are simulated. The two stimuli can also be interpreted as a cross section of a single bar (target) and two flanking parallel bars (mask).

Each iteration of the lateral inhibitory equations represents 30 milliseconds of time, which is approximately the latency of reciprocal lateral inhibition of on-center and off-center cells in the cat lateral geniculate (Singer & Creutzfeldt, 1970). This is not to imply that metacontrast occurs in the geniculate by this mechanism, but only that 30 milliseconds is a physiologically reasonable latency for lateral inhibitory effects in mammals. Metacontrast should be obtained at two iterations, or 60 milliseconds between the target and masking stimulus onsets.

The effect of the disc stimulus on the network is shown in Figure 3 (top), where the stimulus is applied for two iterations of lateral inhibition and is turned off at the dashed vertical line. The curves introduced in the second iteration of stimulation represent Mach bands. The oscillations occurring after stimulation ceases are a function only of the stimulus and the network characteristics; they are just as specific to the disc as the iterations with the stimulus on. Here a distinction made by Stigler (1910), in the paper where he coined the term "metacontrast," is useful. He called the biological results of a physical stimulus during its appearance the "homophotic" stimulus, and those after its disappearance the "metaphotic" stimulus. The network activity before the dashed line represents homophotic stimulation while that after the line represents metaphotic stimulation. Any learning or recognition mechanism associated with this or subsequent cell layers would encounter both homophotic and metaphotic stimuli, and both would be equally correlated with disc stimulation. Though the metaphotic oscillations do not look like the disc, they are specific to it, an example of Shepard and Chipman's (1970) "second-order isomorphism." Because they represent the stimulus after its physical disappearance, they suggest a mechanism for short-term "iconic" storage (Neisser, 1967, Ch. 2). Both the iconic store and

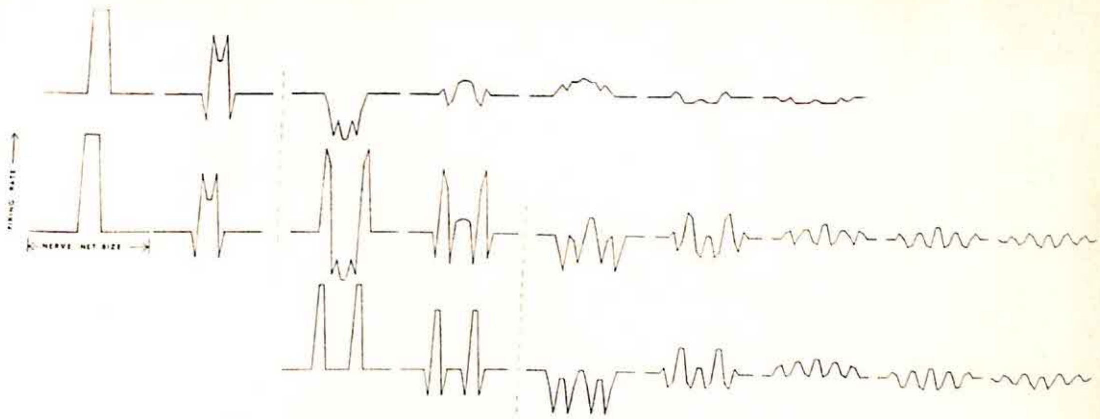


FIG. 3. Simulation of metacontrast. (Top: result of stimulation by the target (disc) stimulus for two iterations. At the dashed vertical line, input is reset to background and the system allowed to oscillate. Middle: metacontrast. The disc is exposed for two iterations. As it is removed, the annulus is applied and also exposed for two iterations. At the second dashed vertical line, all stimulation is reset to background. Stimulus onset difference represents 60 milliseconds of real time. Bottom: stimulation by the mask (annulus) for two iterations. Stimuli are aligned so that direct comparison of effects at corresponding times after stimulus onset is possible. Oscillations after removal of all stimuli in the metacontrast condition are more similar to the annulus than to the disc patterns. Ordinate: firing rate. Abcissa: time.)

the oscillations are modality and stimulus specific, with significant decay in less than one second.

A subject encounters a dilemma when asked to discriminate some feature of a disc presented for a brief time. If the stimulus is present for a shorter time than that needed for perception, the subject must use both the homophotic and the metaphotic stimuli in order to be stimulated long enough for perception to take place. If the metaphotic activity pattern is altered, perception is impaired even while simple detection, taking a shorter time, might still be possible. The curves shown below could be transformed inputs to either a template-matching or a feature-detection perceptual system; no assumptions are required about the perceptual process itself except that it takes time. These ideas can be used to account for both monotonic and U-shaped metacontrast functions with identical stimuli.

A complete metacontrast simulation is presented in Figure 3 (middle), where the disc is first exposed for two iterations and then turned off. At this time the masking annulus is also exposed for two iterations, after which the network is allowed to oscillate with the input set at the background level. For comparison, Figure 3 (top and bottom) also shows the same network stimulated with the disc only and the annulus only at the corresponding times. It is clear that the metaphotic afteroscillations in the metacontrast situation are similar to those generated by the annulus alone, but are very different from the activity resulting from stimulation by the disc alone, so that information specific to the annulus is available in the metacontrast condition, while information specific to the disc is present for too short a time to allow perception to occur.

The situation is quite different when the target and masking stimuli are presented simultaneously as shown in Figure 4, again with disc-only and annulus-only patterns for comparison. The metaphotic stimulus in this case has similarities to both the disc and the annulus conditions, but is distinct from both. One can speculate that the distinctive oscillations give this figure a uniqueness in perception, and therefore in language; it is called not "disc and ring," but "bull's-eye."

Another necessary control condition is the forward masking paradigm, presenting first the masking stimulus and then the target (Figure 5). In this condition, the

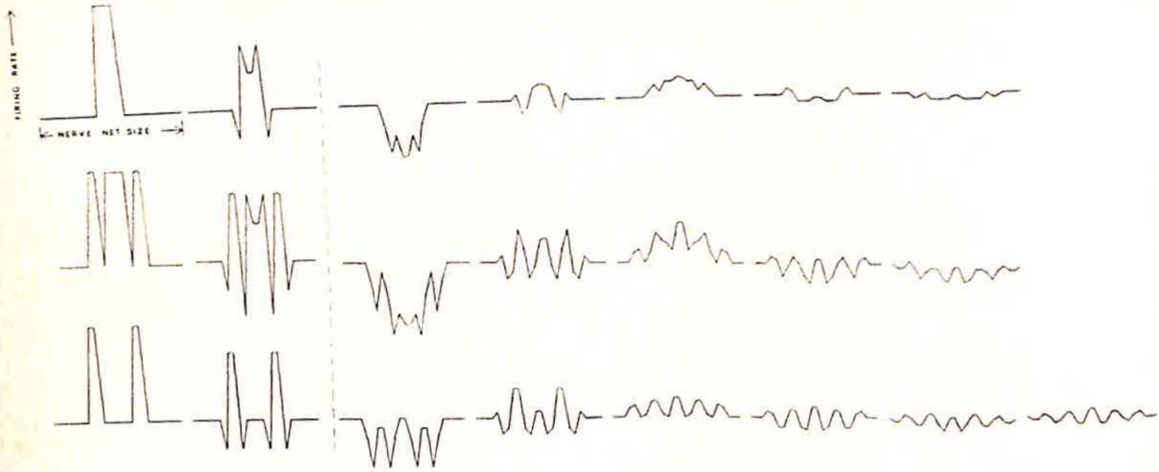


FIG. 4. Simultaneous presentation of target and mask. (Top: disc stimulation only, ending at dashed vertical line. Middle: stimulation with both stimuli simultaneously for two iterations, ending at dashed vertical line. Bottom: annulus stimulation only. Oscillations after stimulus offset for the middle condition are intermediate between those for presentation of either stimulus alone. Ordinate: firing rate. Abcissa: time.)

afteroscillations look at first similar to those generated by the disc alone at the corresponding times after stimulation, but later the annulus effects reassert themselves so that both stimuli should be perceived.

The relationships shown in these results were quantified with the technique of cross-correlation, which was used because it is a way of comparing two curves with a minimum of assumptions (Hays, 1963, p. 510). Each point in Figure 6 represents a cross-correlation of the network's response to a masking condition with the response due to one stimulus alone. The cross-correlation is repeated for each iteration of lateral inhibition, yielding a line which represents the similarity of a masking condition to the disc-only or the annulus-only response over time. Any stimulus array which creates a high correlation with the activity created by the disc, for example, will be recognized by the system as disclike.

Figure 6 shows cross-correlations for all of the curves illustrated in Figures 3–5.

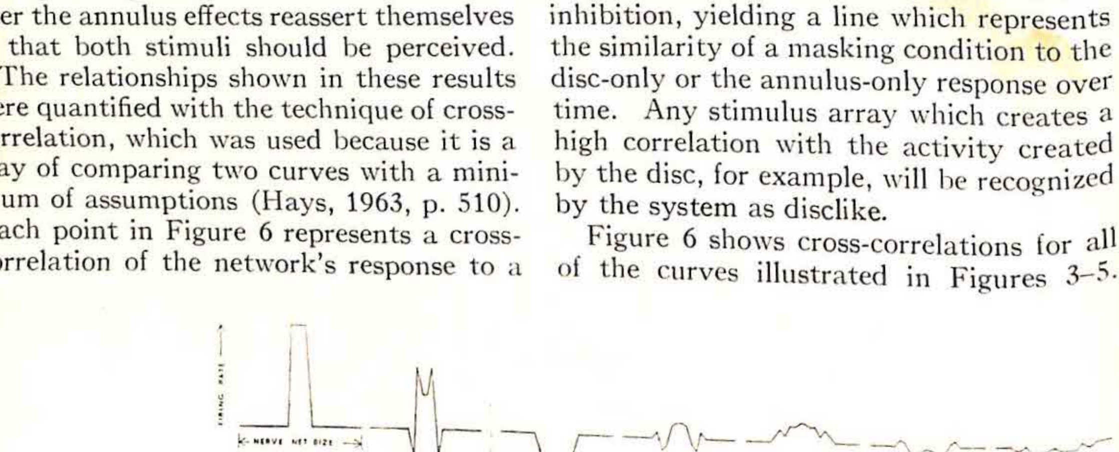


FIG. 5. Forward masking. (Top: disc stimulation only. Middle: stimulation with annulus for two iterations followed by disc for two iterations. Bottom: annulus stimulation only. Oscillations in the masking condition are intermediate between those observed for disc and for annulus-only stimulation. Ordinate: firing rate. Abcissa: time.)

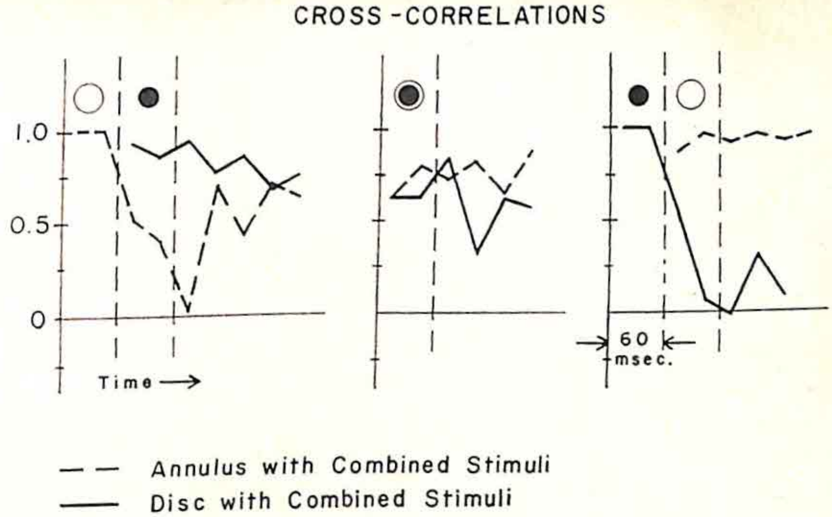


FIG. 6. Cross-correlations. (Each point represents the cross-correlation of the activity generated in the disc-only or the annulus-only condition with that generated in a masking condition. The time axis in each graph is that of Figures 3–5, yielding lines specifying the degree of similarity of single and combined stimuli over time for each condition. Solid lines represent the visibility of the disc. Left: forward masking. Center: simultaneous presentation. Right: backward masking. For the metaphotic period (after stimulus offset), visibility of the disc is minimal in the backward masking condition. For the homophotic period (during stimulation), visibility is minimal for simultaneous target and mask presentation. Ordinate: correlation coefficient. Abcissa: time.)

The left graph shows that the network's response to the control or forward masking condition (Figure 5) correlates fairly well with the response to either of the stimuli alone. The correlation is not perfect for either, of course; this reflects the slight mutual interference seen in all masking experiments, where thresholds for discrimination of both stimuli are raised slightly. Such mutual effects are not explainable by "overtake" theories. The middle graph, for simultaneous disc and annulus presentation (Figure 4), shows somewhat depressed but still highly positive correlations with both single stimuli. Perception can still be quite good in such a situation because of the safety factor built into all biological systems; correlations need not be perfect for recognition to occur. The third graph, however, shows a very different result. For this metacontrast condition (Figure 3), the mask's appearance reduces the correlation of the masking condition with the disc to near zero and holds it there. The disc's metaphotic phase is effectively erased from the network. Thus, if perception requires a fairly high correlation during the time of analysis, perceptual errors will be low for the first two conditions but high for the third. At longer intervals, of course, perception can be completed before the masking stimulus occurs. These effects reflect the **U**-shaped masking function.

For a detection task, with a lower criterion, the result is different. If detection can take place quickly, while the homophotic stimulus prevails, a low-criterion masking can be obtained by examining only the first two iterations of the disc's cross-correlation with the masking condition in each graph (solid lines). These cross-correlations are at a minimum for simultaneous disc and annulus presentation, meaning that detection errors will be maximal here, reflecting the monotonic masking function. Thus both the **U**-shaped and the monotonic masking functions can be obtained from the same patterns of activity, depending on the subject's criterion level.

DISCUSSION

This simulation began with a modified Hartline-Ratliff neural net excited with spatial and temporal patterns like those used in visual masking experiments. Under the appropriate conditions, both U-shaped and monotonic metacontrast functions were obtained, and some mutual masking was also observed. A complete theory of metacontrast, however, must also account for several other important properties of the effect. One of these is the decreasing effectiveness of masking as the angular distance between target and mask increases. In the model simulated here, where interaction effects travel one neuron laterally for each time iteration, effects from two stimuli begin to interact one additional iteration after stimulus onset for each cell interposed between them. Because inhibitory effects are weaker at larger distances, each metaphotic stimulus is less distorted by the other and less masking results. With greater separations, each stimulus has its characteristic oscillations in a different part of the network. In fact the cross-correlation in the brain may be carried out over relatively restricted areas of the visual array. Presence of neural noise would tend to reduce all of the cross-correlations, especially at larger latencies and distances.

Many masking studies find the relative energy of the target and masking stimuli to be of crucial importance. If the mask is very much stronger than the target, monotonic functions are obtained (Fehrer & Smith, 1962; Kolers, 1962). A strong mask reduces cross-correlations of the masking condition with the disc-only condition simply because the high energy of the mask creates a very large disturbance which reduces the target's relative contribution to the total activity. The relative energy of the stimuli, rather than their spatial characteristics, becomes the crucial parameter and the situation becomes similar to that of masking by light flashes.

Though this network can be transposed only with great care from the *Limulus* eye to the human cortex, there is evidence that such a network comprises at least part of the functional anatomy of the cortex (Creutzfeldt, Baumgartner, & Schön, 1956; Krnjevic, Randic, & Straughan, 1964). An additive model of excitation from a receptive field center and inhibition (or off response) from the periphery can account for much of the experimental data on the temporal responses of cat retinal ganglion cells to sinusoidal stimulation (Maffei & Cervetto, 1968; Maffei, Cervetto, & Fiorentini, 1970). Lateral inhibition would create circular receptive fields with an antagonistic surround; even in monkey visual cortex, these comprise about one-quarter of all receptive fields (Spinelli, Pribram, & Bridgeman, 1970). This figure suggests that though a functional lateral inhibitory network may be operating in visual cortex, other modes of neuronal organization are also present. The same authors found that about one-third of the cells they examined could not be influenced by the same simple visual stimuli which could effectively elicit receptive fields in nearly all cells of cat visual cortex and in other monkey cortex cells. These "uncommitted" cells do not fit into the lateral inhibitory scheme and become more common near the foveal projection area. A different sort of processing may be occurring there, more powerful but requiring many more neuronal components. The uncommitted cells may explain why metacontrast is not obtained in the fovea.

Another anatomical alternative for the model is that either lateral or forward inhibition could be applied to incoming cells at several stages of the visual pathway. This is mathematically equivalent to lateral inhibition at the final stage, implying that lateral inhibition in the cortex is not necessary for this scheme to be effective, but only an addition of the inputs from the two eyes. Lateral inhibition is well documented at both retinal (Grüsser et al., 1965; Wuttke & Grüsser, 1968) and lateral geniculate (Singer & Creutzfeldt, 1970) levels.

In addition to accounting for existing

experimental data on metacontrast, this theory makes some predictions. At present there is a lack of good psychophysical data on the nature of the change from monotonic to U-shaped metacontrast functions as response criterion changes. If response criterion is varied systematically in terms of modern decision-theory concepts, the lateral inhibitory model presented here predicts that subjects should switch gradually from monotonic to U-shaped masking functions as their response criterion is increased, with no change in stimulus parameters.

Under normal stimulus conditions, the oscillations of neuronal activity induced by a stimulus damp out to ready the system for new input, while in metacontrast the oscillations of two stimuli interact so that the metaphotic phase of the target stimulus is distorted. The interaction of oscillations can be pushed further, as illustrated in Figure 7. By overdriving a lateral inhibitory network capable of contour enhancement, continual and abnormally large oscillations could be set up. The resulting instability might cause the striking movement effects and aftereffects created by

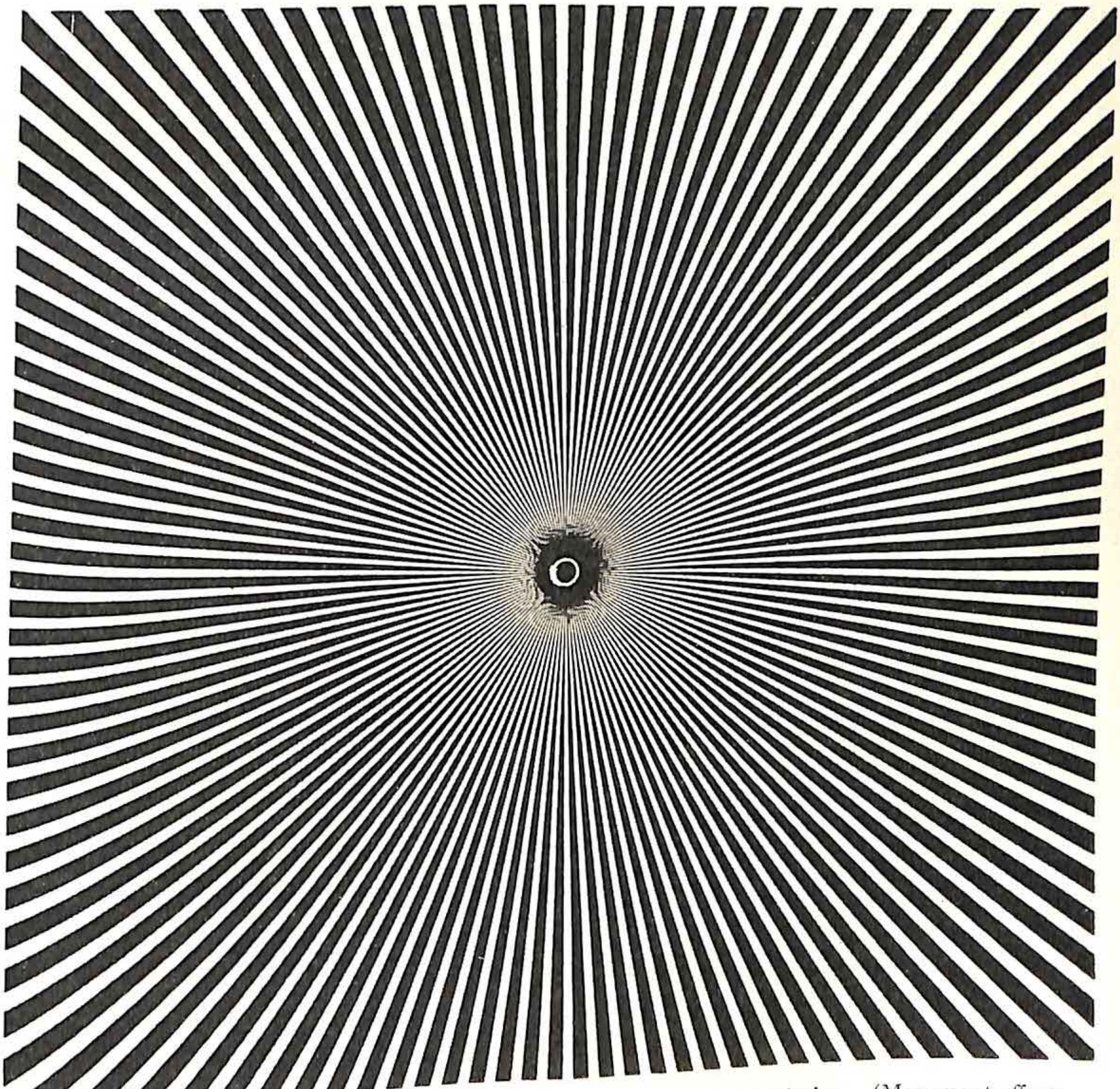


FIG. 7. Result of overdriving the visual system with patterned stimulation. (Movement effects may be due to unstable oscillation of a lateral inhibitory network in the sensory system.)

viewing Figure 7, which hints at the power of the transformations mediated by lateral inhibition. The mathematical requirements for amplification by lateral inhibition have been developed by Ratliff et al. (1969). After processing by the lateral inhibitory network, visual information has some rather special properties; it is no longer isomorphic with the retinal stimulus, and it has been distributed in space and time. Figures 3 through 5 show that the distribution of effects across the network is much wider than the extent of direct anatomical connections. Enhanced spatial distribution increases the reliability of the system because information about the stimulus is now coded in many more elements than the original input neurons, and information readout to another level can be preserved despite failure of readout from any few neurons. While remaining homotopically organized, visual information is spread in such a way that any retinal locus is represented by perturbations over a wide cortical area, and in like manner one cortical locus encodes information about a wide area of the visual world. The extent of this spread is shown by McIlwain's (1964) finding that many single visual cells even in the retina and lateral geniculate of the cat can be influenced by some stimuli as much as 90° from the centers of their receptive fields. This sort of spatial redundancy is suggestive of Pribram's (1969) holographic, or interference-pattern, theory of perceptual and memory coding. On a more microscopic physiological level, output from the lateral inhibitory system would be an ideal input to Spinelli's (1970) OCCAM memory model.

The model proposed here is a mathematical-*neural* model of a substrate whose quantitative properties have not yet been defined in vertebrates, though each parameter has a physiological meaning. The metacontrast effects described above, however, are robust; the entire simulation has been repeated with a single inhibitory coefficient of .4, yielding narrower oscillations but similar masking results. Recent studies by von Békésy (1971) show that metacontrastlike inhibitory phenomena also occur in the auditory system in the same time range as visual metacontrast, about 60 milliseconds between stimuli. He refers to the effect as backward inhibition, a term which is defined behaviorally and should not be confused with lateral inhibition, which is defined anatomically. The present paper supports von Békésy's suggestion that backward inhibition may indeed be mediated by a lateral inhibitory network, and metacontrastlike effects may be found wherever lateral inhibition is used in sensory systems.

The theory applied here to metacontrast was not developed for the purpose of accounting for masking data. The lateral inhibitory equations were originated to explain spatial and temporal interactions observed in *Limulus* and other eyes, and even this very rough model when combined with some psychological data can account for metacontrast phenomena better than existing theories without the postulation of any other neuronal processing. The approach invokes no mechanisms which are not known to be present in the nervous system; metacontrast falls out as an artifact of the lateral inhibitory way of processing information, a price which the organism pays for gaining the power of the network to perform other functions such as intensity adaptation and contour enhancement.

REFERENCES

ALPERN, M. Metacontrast: Historical introduction. *American Journal of Optometry*, 1952, **29**, 631–646.

BARLOW, R. B. Inhibitory fields in the *Limulus* lateral eye. *Journal of General Physiology*, 1969, **54**, 383–396.

CREUTZFELDT, O. D., BAUMGARTNER, G., & SCHÖN, L. Reaktionen einzelner Neurone des sensomotorischen Cortex nach elektrischen Reizen. *Archiv für Psychiatrie und Nervenkrankheiten*, 1956, **194**, 597–619.

DODGE, F. A., SHAPLEY, R. M., & KNIGHT, B. W. Linear systems analysis of the *Limulus* retina. *Behavioral Science*, 1970, **15**, 24–36.

FEHRER, E., & SMITH, E. Effect of luminance ratio on masking. *Perceptual and Motor Skills*, 1962, **14**, 243–253.

GANZ, L. Mechanism of the figural aftereffects. *Psychological Review*, 1966, **73**, 128–150.

GRÜSSER, O.-J., PETERSEN, A., & SASOWSKI, R. Neurophysiologische Grundlagen des Metakontrastes. *Pflügers Archiv für Physiologie*, 1965, **283**, R 50.

Hays, W. L. *Statistics for psychologists.* New York: Holt, Rinehart & Winston, 1963.

Hull, E. Corticofugal influence in the macaque lateral geniculate nucleus. *Vision Research*, 1968, **8**, 1285–1298.

Kahneman, D. Method, findings, and theory in studies of visual masking. *Psychological Bulletin*, 1968, **70**, 404–425.

Kalil, R. E., & Chase, R. Corticofugal influence on activity of lateral geniculate neurons in the cat. *Journal of Neurophysiology*, 1970, **33**, 459–474.

Kolers, P. A. Intensity and contour effects in visual masking. *Vision Research*, 1962, **2**, 277–294.

Knight, B. W., Toyoda, J., & Dodge, F. A. A quantitative description of the dynamics of excitation and inhibition in the eye of *Limulus*. *Journal of General Physiology*, 1970, **56**, 421–437.

Krnjevic, J., Randic, M., & Straughan, D. Cortical inhibition. *Nature*, 1964, **201**, 1294–1296.

Lange, D., Hartline, H. K., & Ratliff, F. Dynamics of lateral inhibition in the compound eye of *Limulus:* II. In C. G. Bernhard (Ed.), *The functional organization of the compound eye.* Oxford: Pergamon Press, 1966.

Maffei, L., & Cervetto, L. Synamic interactions in retinal receptive fields. *Vision Research*, 1968, **8**, 1299–1303.

Maffei, L., Cervetto, L., & Fiorentini, A. Transfer characteristics of excitation and inhibition in cat retinal ganglion cells. *Journal of Neurophysiology*, 1970, **33**, 276–284.

McIlwain, J. T. Receptive fields of optic tract axons and lateral geniculate cells: Peripheral extent and barbiturate sensitivity. *Journal of Neurophysiology*, 1964, **27**, 1154–1173.

Neisser, U. *Cognitive psychology.* New York: Meredith, 1967.

Pribram, K. H. The four R's of remembering. In K. H. Pribram (Ed.), *On the biology of learning.* New York: Harcourt, Brace & World, 1969.

Pribram, K. H. *The languages of the brain: Experimental paradoxes and principles in neuropsychology.* Englewood Cliffs, N. J.: Prentice-Hall, 1971.

Raab, D. H. Backward masking. *Psychological Bulletin*, 1963, **60**, 118–129.

Ratliff, F. *Mach bands: Quantitative studies on neural networks in the retina.* San Francisco: Holden-Day, 1965.

Ratliff, F., Hartline, H. K., & Lange, D. Dynamics of lateral inhibition in the compound eye of *Limulus*. I. In C. G. Bernhard (Ed.), *The functional organization of the compound eye.* Oxford: Pergamon Press, 1966.

Ratliff, F., Hartline, H. K., & Miller, W. H. Spatial and temporal aspects of retinal inhibitory interaction. *Journal of the Optical Society of America*, 1963, **53**, 110–120.

Ratliff, F., Knight, B. W., & Graham, N. On tuning and amplification by lateral inhibition. *Proceedings of the National Academy of Sciences*, 1969, **62**, 733–740.

Ratliff, F., Knight, B. W., & Milkman, N. Superposition of excitatory and inhibitory influences in the retina of *Limulus:* Effect of delayed inhibition. *Proceedings of the National Academy of Sciences*, 1970, **67**, 1558–1564.

Schiller, P. H. Single unit analysis of backward visual masking and metacontrast in the cat lateral geniculate nucleus. *Vision Research*, 1968, **8**, 855–866.

Schiller, P. H., & Smith, M. Monoptic and dichoptic metacontrast. *Perception and Psychophysics*, 1968, **3**, 237–239.

Shepard, R. N., & Chipman, S. Second-order isomorphism of internal representations: Shapes of states. *Cognitive Psychology*, 1970, **1**, 1–17.

Singer, W., & Creutzfeldt, O. D. Reciprocal lateral inhibition of on- and off-center neurones in the lateral geniculate body of cat. *Experimental Brain Research*, 1970, **10**, 311–330.

Spinelli, D. N. OCCAM: A computer model for a content addressable memory in the central nervous system. In K. H. Pribram & D. E. Broadbent (Eds.), *Biology of memory.* New York: Academic Press, 1970.

Spinelli, D. N., Pribram, K. H., & Bridgeman, B. Visual receptive field organization of single units in the visual cortex of monkey. *International Journal of Neuroscience*, 1970, **1**, 67–74.

Stigler, R. Chronophotische Studien über den Umgebungskontrast. *Pflügers Archiv für die gesamte Physiologie*, 1910, **134**, 365–435.

von Békésy, G. Auditory backward inhibition in concert halls. *Science*, 1971, **171**, 529–536.

Weisstein, N. A Rashevsky-Landahl neural net: Simulation of metacontrast. *Psychological Review*, 1968, **75**, 494–521.

Wuttke, W., & Grüsser, O.-J. The conduction velocity of lateral inhibition in the cat's retina. *Pflügers Archiv für Physiologie*, 1968, **304**, 253–257.

(Received April 7, 1971)

Psychological Review
1971, Vol. 78, No. 6, 540–551

TEMPORAL AUDITORY ACUITY[1]

DAVID M. GREEN[2]

University of California, San Diego

Temporal auditory acuity, the shortest time interval within which the ear can discriminate the order of auditory events, is reviewed. Experiments are reported on transient signals in which the task is to discriminate two brief signals that differ only in their phase spectra. Periodic stimuli are next reviewed and it is argued that the same temporal acuity may account for these results. Finally the paper concludes with a discussion of the relation of this work to the research on residue pitch. The review indicates that the value for temporal auditory acuity is about 1 to 2 milliseconds.

The ear's ability to extract information from an acoustic waveform depends on its ability to correctly discriminate among the spectral contents of different waveforms and to correctly discriminate the order in which these waveforms occur. Considerable effort has been directed at studying the ear's capacity to discriminate changes in the spectrum. There are numerous studies of the just discriminable change in frequency Δf or amplitude ΔI of a sinusoid. These are also numerous studies of discrimination of a slight change in a broadband spectrum such as noise. The entire area of masking, studies of how one sound obscures another, can be viewed as investigations of the limits of such spectral discrimination. Very few studies, however, have been directed at determining the ear's capacity to discriminate changes in temporal order, that is, what I will call temporal auditory acuity. Speech and music are but some of the many acoustic waveforms where the temporal order of the acoustic elements is vital in the appreciation of the information content. This paper is devoted to a review of some of the recent efforts to explore temporal auditory acuity and to summarize some findings. Essentially we want to know the limits of temporal resolution. What is the shortest time interval within which the ear can discriminate the order of auditory events?

One of the major technical problems encountered in studies of auditory acuity is that it is difficult to generate two auditory stimuli which differ only in temporal order. Often alteration in the temporal waveform produces a number of other unwanted cues, for example, slight differences in the energy spectrum. These spectral alterations may then become the basis for discrimination between the waveforms rather than the temporal differences. In the work described, the energy spectrum of the stimulus was always held constant; thus it can never serve as the basis for discrimination. The waveform is changed by manipulating the phase spectrum. One, therefore, could describe this work as a study of phase perception. While this is an accurate description, I have not found it particularly useful in understanding the results. In thinking about mechanisms responsible for discriminating two waveforms that differ only in their phase spectra, I find it simpler and more concrete to think of differences in the order of temporal events within the waveform rather than differences in phase spectra.

The basic procedure used to study auditory acuity is analogous to that used to study visual acuity, that is, simply establish some difference in temporal order between two stimuli and then progressively reduce the total duration of the stimuli until the different orders can no longer be perceived. For a concrete example, think of two clicks, one about half the amplitude of the other. If we play them in one order, we hear loud–soft; in the opposite order soft–loud. No matter which order was used, the energy spectra of the pair of clicks are the same; the pairs differ only in phase, that is, the order of events in time. Now let us gradually reduce the time between the clicks until both clicks occur within some very short interval. It was surprising to find that subjects could reliably make such discriminations even when the duration of the two clicks was only 1 or 2 milliseconds (msec.). The author and his collaborators have performed a number of similar experiments to investigate this ability under a variety of circumstances. The limit of temporal resolution was always 1 or 2 msec.

---

[1] This research was supported in part by the National Institutes of Health, United States Public Health Service, and in part by the National Science Foundation. The paper was initially prepared as an invited address sponsored by Division 3 at the meeting of the American Psychological Association, Miami Beach, September 1970. I would like to thank my collaborators, Don Ronken, Barry Leshowitz, James Patterson, and Roy Patterson, with whom various aspects of this work have been pursued in our laboratory for the past two years.

[2] Request for reprints should be sent to David M. Green, Department of Psychology, University of California, San Diego, La Jolla, California 92037.

A general question is whether the same temporal acuity is possible with steady-state as well as transient signals. There are several ways to tackle that question and some preliminary results are presented. In the limit, one can present periodic stimuli and ask if the subject is able to discriminate phase changes among the periodic components. This is a topic closely related to pitch perception or "residue" pitch and is rather remote from the original starting point; nevertheless, it now seems that the mechanism is basically the same and that these different experiments merely explore different aspects of the same basic process.

TEMPORAL INTEGRATION

The majority of the traditional work on temporal parameters of the auditory system concerns integration time. In these experiments, the subject attempts to detect some weak signal either in a background noise or in quiet. The threshold of this weak signal is measured as a function of its duration. Typically one finds that the signal's detectability is the same if the product of signal duration and intensity is held constant (energy constant) at least over some limited range of durations. Generally, this equal energy rule holds for signal durations of up to 100 to 200 msec. Zwislocki (1960), in his theory of temporal integration, has written and summarized a great deal of the information in this area.

It would be parsimonious, but untrue, to believe that since the ear can add up information over a 100-msec. interval, that it cannot distinguish how the information arrives within that interval. Certain simple mechanical analogies might work that way, but the ear is quite different. Hirsh, for example, has established that one needs only about 25 msec. to distinguish differences in starting times between a high- and low-frequency sinusoid. Patterson and Green (1970), using a different psychophysical procedure and a much shorter signal, found that discrimination is possible when the total duration of the transients is only 2 msec. The latter experiments were carefully controlled to rule out spectral differences as being the basis for the discrimination. The discrimination is based on some sort of qualitative information. One of the transients sounded like "tick," the other "tock." Before describing this experiment in more detail, let us consider the general problem of discriminating between short transient signals and how this ability is related to integration processes.

TEMPORAL ACUITY AND THE EAR'S MINIMUM TIME-CONSTANT

The response of any realistic physical system is a function of the present input and some combination of past inputs. If the variation in the input is too rapid, the inertia in the system smooths that input variation and produces an output that is some weighted combination of the recent past inputs. If the system is linear, the output is simply a weighted average of past inputs. One of the simplest linear systems has an output that is a simple exponential weight of past inputs. Recent values of the input are given larger weight, whereas more remote values are given less weight. The parameter of this exponential decay is called the time constant of the device and represents the memory of the system.

More complicated systems may not have simple exponential weighting functions but if they are linear, one can still represent the present output as a weighted average of past inputs. The weighting function is simply more complicated. In all systems, both linear and nonlinear, it is important to know how much of the past affects the present output.

Now consider how these systems will behave when confronted with two clicks. Figure 1 illustrates the output of the system to the soft–loud and loud–soft click pairs. The output is illustrated assuming two quite different time constants. Because of the sluggish response of the long time-constant device, the order in which the clicks arrive would probably not create a noticeable difference in the overall response. On the other hand, the form of the output for the short time-constant device is obviously heavily dependent on the order in which the clicks occur. Incidentally, Ronken (1970) has shown that observers can distinguish the order of two such pairs of clicks down to about 2 msec.

Figure 1 illustrates the general experimental procedure, at least in its conceptual form. By determining the limits of the discrimination of temporal order for brief acoustic sounds, we can determine the minimum time-constant of the auditory system. For inputs shorter than this time, the response of the system is apparently some combination of events of

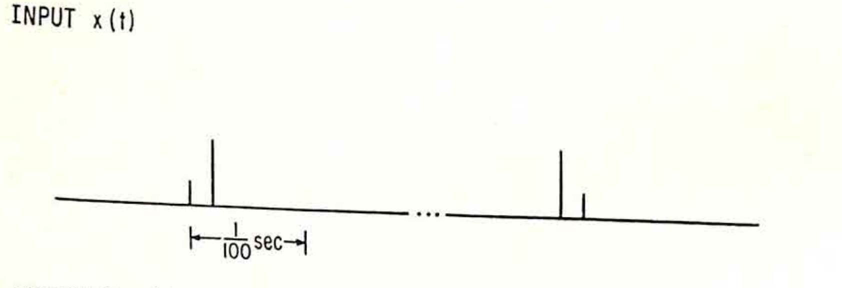

INPUT x(t)
$\frac{1}{100}$ sec
RESPONSE y(t)
LONG TIME CONSTANT $\frac{1}{k} >> 1$ SEC
1 sec
RESPONSE y(t)
SHORT TIME CONSTANT $\frac{1}{k} < 1$ SEC
$\frac{1}{100}$ sec

FIG. 1. Response of two simple linear systems to a pair of clicks. (The top line shows the inputs, only the order of the clicks differ. The middle line shows the response of a system with a long time-constant; the response of the short time-constant device is shown on the bottom line. Note the difference in time scale for the middle line.)

the recent past in which order information is lost.

We must be very careful, however, about details of this experiment when we apply it to the sense of hearing. It is well known that the ear performs a limited spectral analysis and hence we must be very careful to insure that order of arrival is the only cue available to the subject. If the clicks differed in energy spectrum, then the discrimination might be based on spectral differences and the experimental results would then reflect the quality of spectral analysis rather than parameters of temporal processes. To illustrate the importance of this problem, suppose we ask what is the minimum gap between two 10-microsecond clicks that the ear can reliably detect. Leshowitz (1971) has shown that the minimum gap is 6 microseconds and that the basis for the discrimination is undoubtedly slight differences in the energy spectra in the region of 10,000 to 15,000 hertz (Hz.). The detection of the gap is, therefore, not a temporal discrimination but rather a discrimination of difference in intensity at certain regions of the spectrum.

The two-click procedure illustrated in Figure 1 avoids this problem, since the energy spectrum is the same for each pair; the only difference between the pairs is their phase spectrum. What we would like is a large set of stimuli, all with equal energy spectra, that differ only in their phase spectra. In particular, we would like to be able to program how the energy is ordered over the entire frequency range so that we could delay one or another part at will. Huffman (1962) has provided a general means of accomplishing this requirement. He has provided a method for constructing a large set of waveforms with the following important properties: (*a*) the waveforms are zero outside an interval T; (*b*) the energy spectra of all waveforms are the same, only the phase spectra differ. Thus all the waveforms in a class generate the same energy at any frequency; it is only the order in which this energy occurs at various frequencies that differs over the class. We call a single waveform of this class a Huffman sequence.

EXPERIMENTS WITH HUFFMAN SEQUENCES

Figure 2 shows two 10-msec. signals, both Huffman sequences and thus having the same energy spectra. The upper trace is the output of the earphone as measured into a 6-cubic-centimeter coupler; the lower trace is the electrical voltage applied to the earphones. The phase spectra of the two signals differ. The left waveform has a phase delay at 200 Hz.; the right waveform has a phase delay at 1000 Hz.

The relative energy and phase spectra of these signals are shown in Figure 3. The energy spectra are identical for both signals and are marked "magnitude" in the figure. The falloff in magnitude is caused by quantizing the waveform in time. For all of the experiments, the signal was filtered to exclude this high frequency energy (marked in grey on the figure); thus all the signals will have an approximately rectangular spectrum. The phase spectra are shown below in the bottom portion of the figure. The difference in the phase spectra means that the energy associated with different frequencies arrives at different times. The ability to distinguish among these waveforms means the ear is sensitive to these differences in the arrival time of energies at different frequencies.

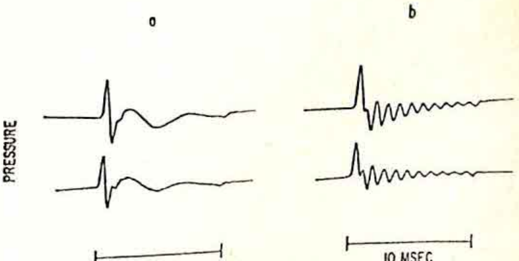


FIG. 2. Two 10-msec. Huffman waveforms. (Part a shows the waveform generated by the 200-Hz. Huffman sequence filter with a phase delay near 200 Hz. The upper trace is the output of the earphone into a 6-cubic-centimeter coupler. The lower trace is the voltage waveform applied to the earphones. Part b shows another Huffman sequence generated by phase delay near 1000 Hz.)

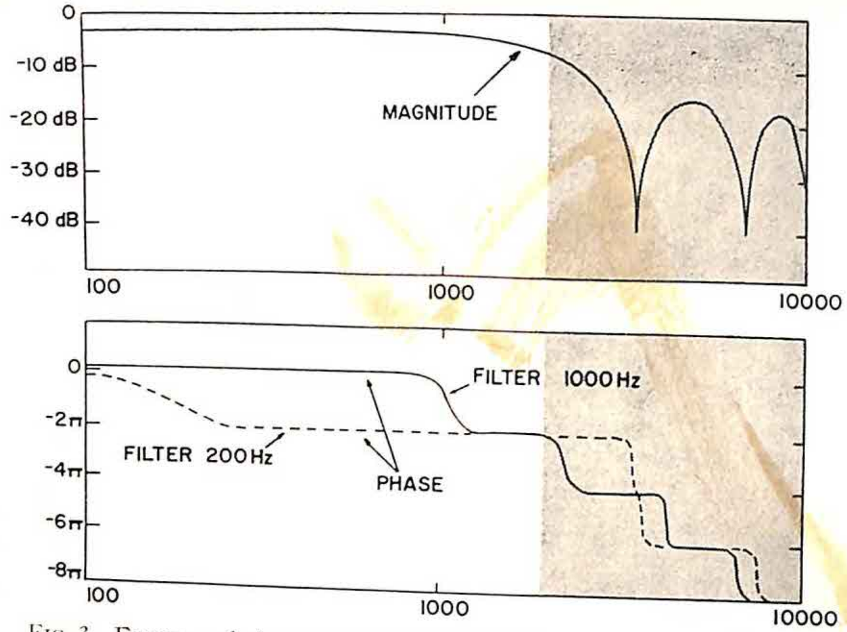


FIG. 3. Energy and phase spectrum of two Huffman sequences. (The curve, labeled Magnitude, gives the energy spectrum in decibels (arbitrary reference) of Huffman sequences generated at 3200 samples per second. The gray area indicates that analog filtering was used to suppress all frequencies above half the sample rate. Thus, the magnitude spectrum is essentially flat. The phase curves of two different sequences are shown below, one with its singularity at 200 Hz., the other with its singularity at 1000 Hz. The two phase curves correspond to waveforms a and b of the previous figure.)

Consider a short time-power spectrum of these signals such as that provided by a speech spectrogram. In the conventional speech spectrogram, energy is coded by the darkness of the marking, and the coordinates of the graph are frequency (along the *y* axis) and time (along the *x* axis). Thus, a single sinusoid is represented by a horizontal line at the appropriate frequency and a brief burst of white noise by a dark vertical line. If a typical speech spectrogram were used to analyze the signals, then they would all be nearly identical, namely, thick dark vertical bars. This is true because the different spectra are nearly white and the signals are brief compared with the time constant of the speech analyzers, which is about 20–30 msec. If the analyzer is modified by altering the time constant by changing it, say from 20 to 1 msec., then differences in the signals would become more evident. With these brief time constants, it would be clearly seen that the signals differ in the arrival time of energy at the different frequencies. For example, Figure 4 shows a short time-spectral analysis of the two waveforms displayed in Figure 2. The large spike at the beginning of the time waveform corresponds to the large vertical line at the left of the spectrogram. The near simultaneous arrival of energy over a wide range of frequencies causes the large spike in the time domain. The energy at one frequency, about 200 Hz. in Figure 4a and about 1000 Hz. in Figure 4b is delayed. If we listened to the two waveforms, both would sound like a click, but the waveform with the energy delayed at 1000 Hz. would sound higher in pitch, "tick," than the waveform with delayed energy at 200 Hz. "tock."

The first experiments consisted of simply shortening the total duration of these signals until they could no longer be discriminated in a paired-comparison test. Subjects can achieve about 100% discrimination on such waveforms if they are 10 msec. in duration, they achieve near chance at 1 msec., and they are correct

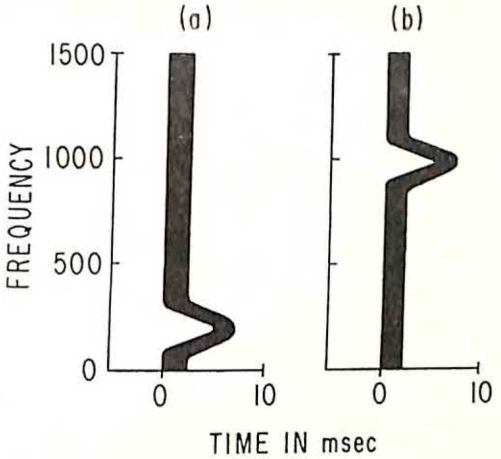


Fig. 4. Time-frequency distribution of energy for two waveforms. (The figure shows the locus of energy as a function of time and frequency for two Huffman sequences. The (a) graph is the distribution for a Huffman sequence generated with a single phase delay at 200 Hz., e.g., the waveform marked a in Figure 2. The (b) graph corresponds to waveform b of Figure 2. It has a phase delay at 1000 Hz.)

about 75% of the time when the duration is about 2 or 3 msec. The ability to distinguish between pairs of Huffman sequences also depends on other variables such as the frequency separation between the phase delays and the width, in frequency, of delayed energy. The value of 2 msec. is an upper bound on the true temporal acuity value because that number is the total duration of the signal. Distinctions among the waveforms must be based on differences that occur within this total duration.

In addition to experiments on discriminating between pairs of such signals, Patterson and Green (1970) have also investigated other properties of these signals. Since these transient signals have broadband, noiselike spectra, the question of how well they mask sinusoidal signals was considered. In these masking experiments, the shape of the envelope of the waveforms within the groups was systematically varied. At one extreme, the envelope was nearly uniform over the 10-msec. duration. At the other extreme, the envelope peaked near the onset and decayed rapidly. Thus, in the latter case, most of the masker energy was concentrated in the first few milliseconds of the observation interval. Three groups of maskers with envelopes intermediate between these extremes were also used. It should be kept in mind that all the waveforms in all groups had identical energy spectra. Thus, if masking is strongly influenced by the energy spectrum of the masker, we might expect all groups of maskers to mask sinusoidal signals equally well. This is clearly not the case. Figure 5 shows the threshold of a 1500-Hz., 10-msec. signal for each of the five groups of maskers. The solid line connects the averages of the four observers. The abscissa of Figure 5 is marked "the number of phase delays"; this scale is an indication of the peak factor in the waveform. If only a few delays are used, the waveform has a high peak factor, such as Figure 2 which has only one phase delay. As the number of phase delays increases, the waveform is nearly uniform over the interval; 25 delays out of 50 possible delays has a very uniform envelope pattern. As can be seen, the maskers with the more uniform envelope produce considerably more masking than the waveform with the sharp spikes. Plotted on the edges of Figure 5 are the logical extremes of this continuum. To the left is the masking produced by a very brief click at the beginning of the 10-msec. interval. Such a masker produces little masking, although its power spectrum is similar to that produced by the Huffman signals. At the extreme right is the masking produced by random noise having the same average power spectrum as the Huffman sequences.

The fact that a brief click and a random noise sample fall in a natural continuation of the results obtained with the Huffman sequences indicates that the crucial masking variable is the ability of the masking to cover the interval of the signal nearly evenly with energy. Signals that obscure only part of the entire signal duration are not effective maskers. This result also indicates that the ear is very quick and can take advantage of these brief time gaps, even when the total signal duration is only 10 msec.

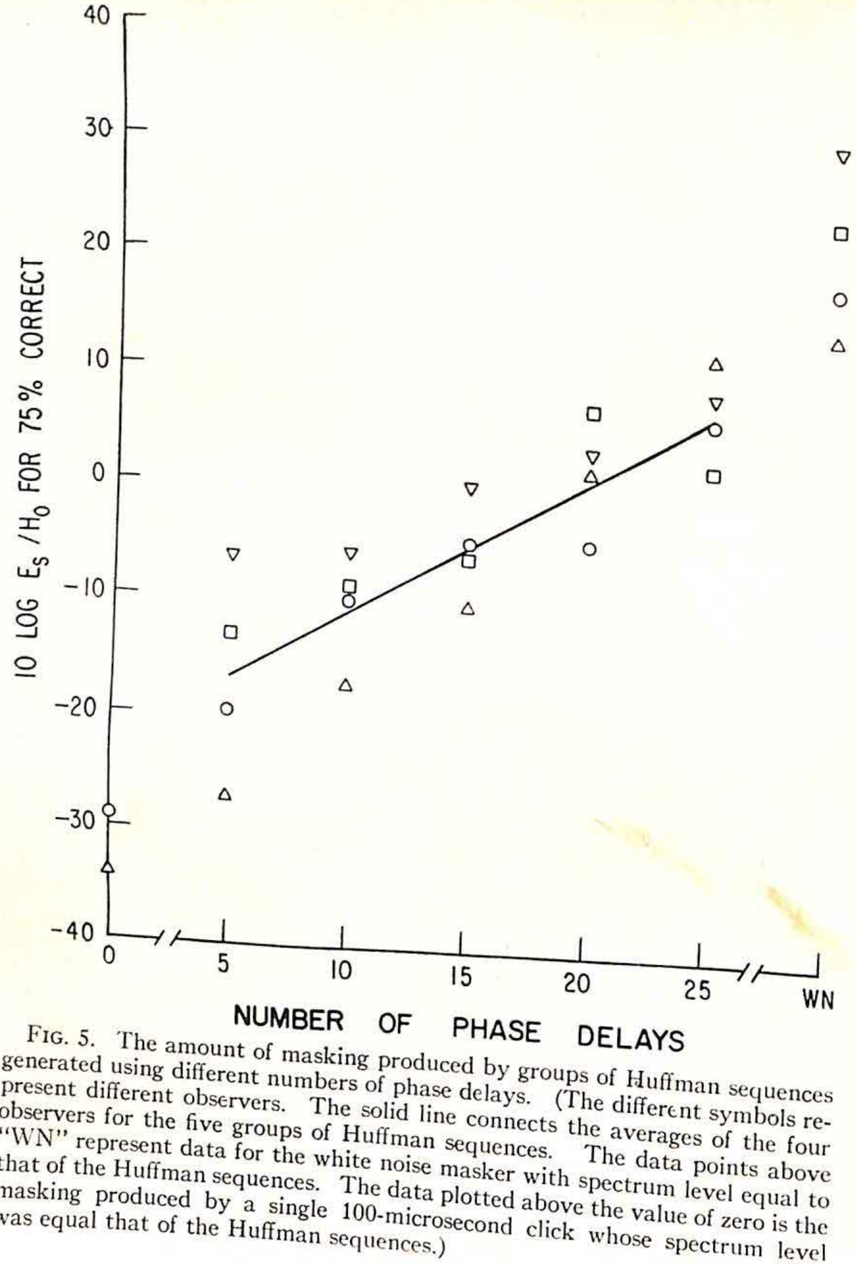


FIG. 5. The amount of masking produced by groups of Huffman sequences generated using different numbers of phase delays. (The different symbols represent different observers. The solid line connects the averages of the four observers for the five groups of Huffman sequences. The data points above "WN" represent data for the white noise masker with spectrum level equal to that of the Huffman sequences. The data plotted above the value of zero is the masking produced by a single 100-microsecond click whose spectrum level was equal that of the Huffman sequences.)

While these results suggest significant differences among Huffman sequences, there is one situation in which the various waveforms within the same class behave quite similarly. This is the threshold for detecting the presence of one of these transient signals in broadband noise. All the Huffman sequences within the same class, that is having identical energy spectra and duration, have very similar masked thresholds. Because they all have similar detection thresholds, one can also determine how much one must increase the signal before it is possible to distinguish one Huffman sequence from another. Figure 6 shows both the detection and discrimination data of two Huffman sequences on a single graph. The curve at the left is the detection data. The triangles are one signal, the circles the other. The curve at the right shows the discrimination data of these same signals; the data points are squares. Once the signals are only 6 decibels above their

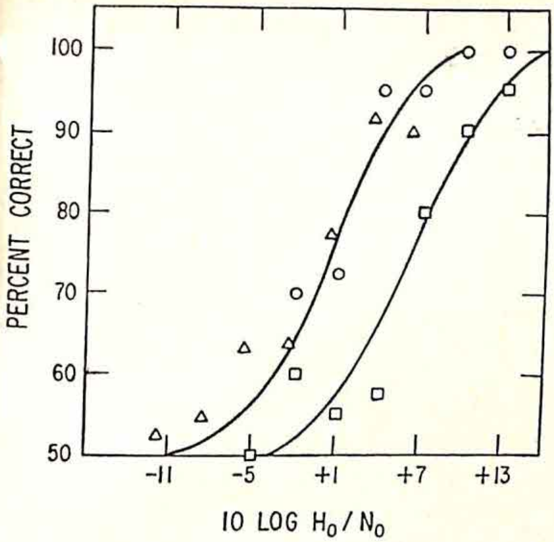


FIG. 6. Percentage of correct discriminations and detections as a function of signal-to-noise ratio. (The squares are percent correct *discriminations* of a pair of waveforms with phase delays either at 400 Hz. or 600 Hz. The circles are percent correct *detections* for the waveform with a 400-Hz. singularity. The triangles are percent correct *detections* for the waveform with singularity at 600 Hz.)

threshold value, it is possible to distinguish one sequence from another. Thus, the ability to distinguish Huffman sequences, although based on slight qualitative differences in the sounds of the two transients, is relatively robust and not easily masked.

Furthermore, Ronken (1970) has shown that even when phase changes are an irrelevant cue in detecting a change in amplitude of two brief clicks, the randomization of this cue interferes with the discrimination of the intensity change indicating that this phase cue is difficult to ignore.

Thus, to briefly summarize the results to this point, it has been established that (*a*) the ear can distinguish temporal events differing in time by only a few milliseconds, (*b*) this phase discrimination is not easily masked by noise, (*c*) differences in the phase spectrum can influence the detection of a sinusoid by as much as 30 decibels, and (*d*) phase differences are of sufficient saliency to disrupt amplitude discrimination when phase is a random but irrelevant cue.

But what of the generality of these findings? Perhaps the ear shows this remarkable temporal acuity only when listening to isolated, transient stimuli. How does this ability stand up when the acoustic events are not transient clicks but ongoing sustained sounds ?

TEMPORAL ACUITY WITH LONGER SIGNALS

Work in this area is not complete, but all the preliminary results indicated that the ear's temporal acuity is the same for sustained sound as it is for transients, and indeed this generalization holds for steady-state signals such as periodic tones.

Leshowitz has determined that this remarkable temporal acuity is true for longer duration stimuli such as speechlike sounds. He used synthetic fricatives which are essentially noise bursts having energy throughout most of the audio spectrum. In normal speech, these bursts are generally about 20–50 msec. in duration. For his experiment, he selected stimuli of 50-msec. duration. Using a digital filter, he subjected these noise bursts to a pair of filtering operations. The first filtering was similar to that imposed by the acoustic resonance of the vocal system. The resulting waveform will be called a normal fricative. The second filtering was similar to the first except the impulse response was reversed in time. This produced another waveform, which will be called an abnormal fricative. The abnormal and normal fricatives are similar in many ways but *differ* because of local time reversals. It is something like taking a tape recording and reversing small segments of the tape. The intensity patterns within both bursts remain the same and the distribution of energy over frequency is the same in both sounds; however, locally, time is reversed.

Technically, the operation maintains the same overall energy spectrum for the two noise bursts and thus discrimination on the basis of long-term energy cues is impossible. Discrimination must be based on differences in the temporal order within the stimuli. Using noise bursts of up to 50 msec. in duration, Leshowitz's subjects can easily discriminate among them when the local time reversals are only 5 msec. Within a 5-msec. interval, the observer can apparently discriminate the order of events and reliably distinguish between two waveforms that differ only in this manner.

TEMPORAL ACUITY OF PERIODIC STIMULI

The easiest way to approach this subject is to return for a moment to Huffman sequences. Suppose we play a 10-msec. Huffman signal repetitively, butting the end of one presentation against the beginning of the next. Obviously the waveform repeats 100 times per second and the resultant sounds like a complex, musical chord. The spectrum consists of the fundamental at 100 Hz. and all the harmonics of that frequency. At the component associated with the phase shift, for example, at 1000 Hz. for the signal on th right in Figure 2, there is a phase delay and the energy at that frequency occurs i the middle of the waveform. If one listens to such a complex waveform, one hears a low-frequency buzz; the pitch is that of a pulse train at 100 Hz., and in addition, another high-frequency pitch. The frequency of this high-frequency pitch corresponds to the out-of-phase component, in this case the 1000-Hz. component.

The first to observe this general phenomenon was Manfred Schroeder, who utilized this effect on a special organ at the Bell Telephone Laboratories about 10 years ago. Schroeder rotated the phase of the higher components of a complex periodic wave and was thereby able to play simple tunes. His work is mentioned briefly in Pierce (1960).

There is nothing especially mysterious about this effect. Its existence can be

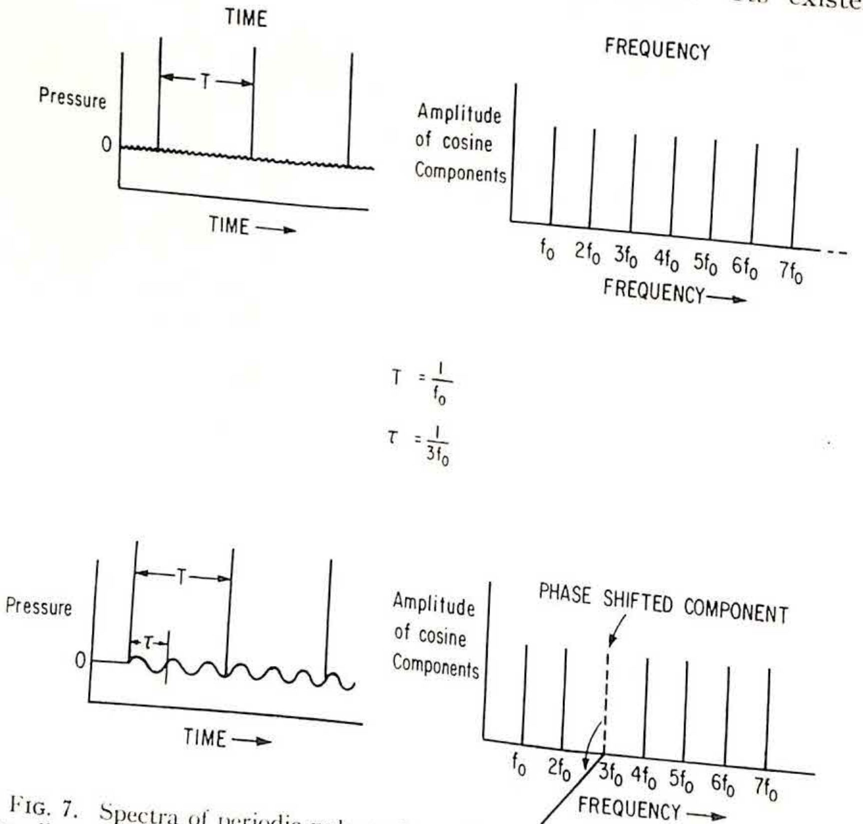

TIME
Pressure
0
T
TIME →
FREQUENCY
Amplitude of cosine Components
f0 2f0 3f0 4f0 5f0 6f0 7f0
FREQUENCY →
T = 1/f0
τ = 1/3f0
Pressure
0
T
τ
TIME →
Amplitude of cosine Components
PHASE SHIFTED COMPONENT
f0 2f0 3f0 4f0 5f0 6f0 7f0
FREQUENCY →

FIG. 7. Spectra of periodic pulse trains. (The waveform and spectra corresponding to a periodic wave in which all components are in cosine phase (top) and that same complex wave with one component ($3f_0$) shifted to a sine rather than cosine phase.)

deduced by the following thought experiment. Suppose we arrange a set of cosine components, a fundamental, and all its harmonics, all having equal amplitude. If all the components are in cosine phase, then we produce a waveform such as shown in Figure 7 (top). The spectrum corresponding to that waveform is illustrated in the right of the figure. If we look at the waveform in time, we find the coherence of all the cosine waves produces the large spike every T seconds. Between the spikes, the various components nearly cancel each other and little energy is present. The distance between the spikes, T, is the reciprocal of the fundamental of the spike train.

Now consider the same set of cosine components, except one of the components has been phase shifted to a sine rather than cosine wave. In Figure 7 (bottom right), we illustrate the third component as rotated 90°. The waveform corresponding to this case is shown at the bottom left in Figure 7. The waveform is little changed from the previous, all-cosine wave. Since one cosine wave is not present, the peak of the spike is reduced by one over the total number of components, and the phase shifted component ($3f_0$) now appears as a sine wave in the time between spike trains. It is easy to see that at some fundamental, f_0, one must be able to hear the phase shifted component. In the limit as the fundamental is moved to lower and lower frequencies, the time between spikes increases and eventually the presence of the out-of-phase component should be evident. In effect, we are studying the audibility of a sinusoidal component masked by two bracketing clicks. If the clicks are separated by sufficient time, one second for example, then any masking produced by the clicks should decrease to such an extent that the phase-shifted sinusoid would be evident.

Thus, the central question is a quantitative one: Under what conditions, what fundamental frequency, what phase shift, and what frequency of the phase-shifted component will the odd component be perceptible? Again it is obvious that we can view this experiment as a measure of temporal acuity. How quickly can the ear recover from the effects of the clicks and discriminate details of the acoustic waveform between the clicks? Systematic measurements have recently been reported on this question in a brilliant paper by Duifhuis (1970). His analysis of these experiments is consistent with the preceding arguments and he emphasizes how these measurements can be used to estimate the quality of the frequency analysis that precedes the temporal analysis. His later data (Duifhuis, 1971) suggest that the temporal resolution at low frequencies (500 Hz.) may not be as acute as that possible at high frequencies (2000 Hz.).

It is also of interest to consider an amusing variation of this experiment. Suppose all components are exactly in cosine phase and the waveform is simply a series of spikes at the reciprocal of the fundamental frequency as shown at the top of Figure 7. If we decrease the amplitude of one of the components, then we are, in effect, adding that component to the pulse train exactly out of phase. Eventually, as we decrease the component sufficiently, the sinusoidal oscillation between the pulses reaches a level that we can hear. We are turning down a component so we can hear it. Thus, one can study both the threshold for presence of a component and the threshold for the absence of a component within the same experimental context.

Many parts of the preceding analysis have been rather superficial. The author has been referring to pictures of sinusoidal oscillations between two clicks and claiming that if you can see the oscillations then you must be able to hear them. Obviously the interest is in the exact parameters required to hear the phase-shifted components, and such measurements are presently being made. It is fairly easy to show that the ear is more complicated than the simple arguments concerning these pictures would indicate. As we know, a limited frequency analysis of the waveform is performed mechanically in the cochlea. Thus, it is possible to hear phase-shifted components

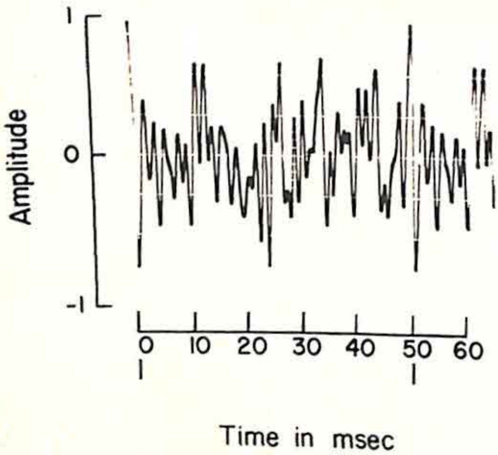


FIG. 8. A periodic wave with considerable phase scrambling. (A periodic waveform in which many components are at random phase but within a block of several cosine phase components, one component is odd (sine phase). While not apparent from the waveform, the odd component is clearly audible.)

when their presence is far from obvious in a wide-band picture of the waveform. Figure 8 shows such a case.

The waveform shown in Figure 8 is a periodic wave in which the phases of many of the components have been scrambled, yet over a certain range (1.4 to 2.8 kilohertz), all of the components are coherent except for one (2.0 kilohertz). The pitch of this one odd component is still evident if one listens to the waveform, but its presence is not clear in a wideband picture of the waveform. By studying the range of components that must be coherent in order to make the odd component still detectable, we will gain further information on the degree of frequency analysis that precedes the temporal processes that have been the main topic of this paper.

TEMPORAL ACUITY AND THE RESIDUE

Throughout our studies of temporal acuity, I have been impressed that the limits on temporal discrimination and the limits on the perception of the residue are very similar. For those not in audition, I might simply explain that the residue is a pitch sensation presumably based on some temporal features of the waveform. It is often illustrated by playing a complex pulse train such as that illustrated in Figure 7. One can then filter out all the lower frequency components of the complex wave and still the low-frequency "pitch" remains. The existence of this residue pitch is frequency cited as evidence against a purely resonance or place theory of pitch perception. The range of the residue is limited; it becomes more and more difficult to hear as the frequency is increased. The upper frequency limit is usually claimed to be about 1000 Hz.

The most direct estimates of this upper frequency limit have been carried out using an experimental procedure first employed by Miller and Taylor (1948). They interrupted white noise at various rates and determined if subjects could set an oscillator at the frequency of the interruption. Since the long-term spectrum of the noise is flat, whether interrupted or not, the basis of the subject judgments must be temporal rather than spectral characteristics of the stimulus. Miller and Taylor found that subjects would make reliable judgments of the repetition rate up to 500–1000 Hz. Several replications of this experiment have been carried out with the same basic result (Harris, 1963; Pollack, 1969), although some subjects seemed to make accurate pitch matches up to interruption rates of 2000 Hz.

It seems likely that the studies of temporal acuity are, in effect, studies of how quickly the peripheral nervous system can respond to changes in the auditory waveforms. These temporal limitations set an upper bound on the frequency that can be utilized by the temporal pitch system. As long as the time T between epochs in the periodic wave is larger than 1 to 2 msec. (500 to 1000 Hz.), then the ear can follow details of the waveform and a residue pitch can be heard. For waveforms with smaller periods, that is, higher frequencies, the ear cannot follow details of the waveform and is forced to utilize spectral place information to determine pitch.

A series of related experiments dealing with auditory temporal acuity has been described. This series of studies deals with the shortest times during which the ear can distinguish the order of temporal events. In all these experiments, the spectral energy distribution, computed over some longer period of time, cannot be the basis for the discrimination. The summary is quite simple. The ear can resolve events within 1 or 2 msec. This resolution is true for transients, for brief 50-msec. noise bursts, and for continuous periodic stimuli. This temporal parameter apparently imposes an upper limit on the frequency that can be utilized by the temporal pitch mechanism.

REFERENCES

DUIFHUIS, H. Audibility of high harmonics in a periodic pulse. *Journal of the Acoustical Society of America*, 1970, **48**, 888–893.

DUIFHUIS, H. Audibility of high harmonics in a periodic pulse: II. Time effects. *Journal of the Acoustical Society of America*, 1971, **49**, 1155–1162.

HARRIS, G. G. Periodicity perception by using gated noise. *Journal of the Acoustical Society of America*, 1963, **48**, 1129–1233.

HUFFMAN, D. A. The generation of impulse-equivalent pulse trains. *IRE Transactions*, 1962, IT 8, S10–S16.

LESHOWITZ, B. The measurement of the two-click threshold. *Journal of the Acoustical Society of America*, 1971, **49**, 462–466.

MILLER, G. A., & TAYLOR, W. G. The perception of repeated bursts of noise. *Journal of the Acoustical Society of America*, 1948, **20**, 171–182.

PATTERSON, J., & GREEN, D. M. Discrimination of transient signals having identical energy spectra. *Journal of the Acoustical Society of America*, 1970, **48**, 894–905.

PIERCE, J. R. Some work on hearing. *American Scientist*, 1960, **48**, 40–45.

POLLACK, I. Periodicity pitch for interrupted white noise—fact or artifact? *Journal of the Acoustical Society of America*, 1969, **45**, 237–238.

RONKEN, D. Monaural detection of a phase difference between clicks. *Journal of the Acoustical Society of America*, 1970, **47**, 1091–1099.

ZWISLOCKI, J. J. Theory of temporal auditory summation. *Journal of the Acoustical Society of America*, 1960, **32**, 1046–1060.

(Received May 14, 1971)

Psychological Review
1971, Vol. 78, No. 6, 552

THEORETICAL NOTES

RANDOMIZATION AND THE SHAPE FUNCTION MODEL OF LEARNING:

A REPLY TO WIESEN

D. M. WARBURTON[1]

Reading University, England

A reader of the Wiesen correction of an error in a paper by Warburton and Greeno is likely to form a gross misconception of the factors motivating the shape function analysis and of the problem of randomization in the construction of small groups.

Weisen (1971) is certainly correct in pointing out an error in the calculation of the probabilities in our example illustrating the dangers of blind randomization (Warburton & Greeno, 1970), but he himself is wrong in his combinatorial arithmetic, in his views on the justification of the shape function model, and in his ideas on randomization in general.

The correct answer to our example is $p = .01$ (actually .00994), the probability of obtaining 7 of one kind from a pool of 16 subjects, consisting of equal numbers of "bright" and "dull" rats. This value can be derived from the general formula of the hypergeometric distribution which is appropriate for situations where the sampling is done without replacement. Weisen's calculations were based on a sampling with replacement model which just shows how easy it is to misapply elementary probability theory, to use Weisen's words.

Wiesen also creates a false impression that the shape function model was justified by this erroneous example alone, whereas the problem of randomization in small group designs was a fairly minor consideration motivating the shape function analysis. The major justification was provided by the comparison of the traditional statistics used to describe acquisition performance and the more precise shape function analysis. The increased precision in description prevented the occurrence of both Type I and Type II errors and the resulting erroneo interpretation of treatment effects.

Finally, he is mistaken in infering that argue against using randomization. Randon zation is useful for increasing accuracy by moving unsuspected biases and, of course, f making one's data reasonably convincing t other people as well as oneself (Cox, 1962 However, these objectives can only be ach to if the experimental arrangements produc up randomization are unsystematic. On this po Cox (1962) stated "randomization is very ver, in large and moderate-sized experiments, the not much good in very small single exper ces- [p. 90]." However, the shape function me can help to guard against any unusual asp e of the experimental arrangements even in s to experiments and also provides a means als recting for the subject's initial response Obviously, if there were marked unexplai differences in the initial response level of o group of subjects, the inferences drawn fro the corrected data would have to be interprete cautiously.

REFERENCES

Cox, D. R. Discussion. In L. J. Savage et al., *The foundations of statistical inference.* New York: Wiley, 1962.

Warburton, D. M., & Greeno, J. G. General shape function model of learning with applications in psychobiology. *Psychological Review,* 1970, 77, 348–352.

Wiesen, J. P. Comment on "General shape function model of learning with applications in psychobiology." *Psychological Review,* 1971, **78,** 272–273.

(Received March 3, 1971)

---

[1] Grateful acknowledgment is given to Bernard Singer and Michel Treisman for their important suggestions during the preparation of this paper.

Requests for reprints should be sent to D. M. Warburton, Department of Psychology, University of Reading, Building 3, Earley Gate, Whiteknights, Reading RG6 2AL, England.

Psychological Review
1971, Vol. 78, No. 6, 553–555

EBBINGHAUS' DERIVED-LIST EXPERIMENTS RECONSIDERED[1]

WAYNE SHEBILSKE AND SHELDON M. EBENHOLTZ[2]

University of Wisconsin, Madison

The results obtained by H. Ebbinghaus in his derived-list paradigm are at odds with inferences from interference theory and are contrary to results obtained in modern derived-list experiments. This paper challenged the contention of D. T. Hakes and R. K. Young that Ebbinghaus used an inappropriate control condition as well as their conclusion that his results might represent various degrees of negative transfer. N. T. Slamecka's contention that Ebbinghaus' results were due to his knoweldge of the transformations also was challenged. Possible causes of this unresolved discrepant finding were discussed in the context of the unique features of Ebbinghaus' procedure.

Ebbinhgaus (1913) devised the derived-list paradigm to test two alternative explanations of serial learning. Both hypotheses explained serial learning in terms of interitem associations, but, according to one, associations are formed only between adjacent items with the possible connection between nonadjacent items vented by the presence of intermediate ms and, according to the other, all the items bound together, but the strength of the ociation is a decreasing function of the ber of intervening items. In his experi-ts, Ebbinghaus learned a first list and 24 rs later a second, derived list. The latter formed so that its adjacent syllables had separated in the first list by either 1, 2, r 7 intervening members. Ebbinghaus soned that if no associative strength was uilt up between nonadjacent syllables, then no savings would result in the derived list, and if such associations were present, then savings would result. He did not entertain a conception of associative interference, that is, that having the old association might inhibit the learning of a new association in the derived list. Given our contemporary view of interference, as Hakes and Young (1966) have noted, the first-list associations would be a source of powerful interference in learning the derived list, and therefore both the theory of direct associations and the theory of remote associations would predict negative transfer for the derived-list paradigm. Although more recent studies using derived lists have found negative transfer (Hakes, James, & Young, 1964; Young, Hakes, & Hicks, 1965), the fact is that Ebbinghaus obtained positive transfer. Using the difference in time to learn[3] the original and derived lists, he reported savings, averaged over two experiments, of 137, 89, 73, and 42 seconds for derived lists with 1, 2, 3, and 7 intermediate syllables skipped, and a savings of only 6 seconds for a specially chosen control list designed to determine the amount of savings attributable to increased item familiarity.

Hakes and Young (1966) attempted to explain this apparent inconsistency as follows:

> Since Irion (1946) has shown that the random-order condition (Ebbinghaus' control condition) yields *negative transfer*, Ebbinghaus' findings cannot be considered to show positive transfer: The savings in the derived-list conditions were greater than for a negative-transfer condition and might well represent decreased negative transfer rather than positive transfer [p. 249].

The authors argued that whether positive transfer was obtained by Ebbinghaus is indeterminate since Ebbinghaus did not use the appropriate control. The implicit assumption here is that if Ebbinghaus had compared his derived-list conditions to a neutral C-D control list, that is a second list composed of new items instead of the original list, he would have found negative transfer as Young et al. (1965) did when they used the C-D control. The C-D list is necessary to control for warm-up and learning-to-learn effects when naive subjects are used. However, one cannot assume that Ebbinghaus, who learned many lists per day, benefited from either of these nonspecific transfer effects (Postman, 1968). For Ebbinghaus, a new C-D control list and the initial list learned 24 hours apart probably would have been learned in the same amount of time, on the average. Therefore, Ebbinghaus' measure of savings was appropriate and his report of positive transfer in derived lists stands as valid. The discrepancy between Ebbinghaus' and Irion's results which was referred to by Hakes and Young (1966) might easily be explained by noting that Irion's random-order condition was *not* the same as Ebbinghaus' control. Irion's random list was a random rearrangement of the members of the first list. Ebbinghaus' random control, on the other hand, was not a reshuffling of any one list, but each randomly derived list was made up of items which had previously been members of *different* original lists. Ebbinghaus learned six 16-item lists for each type of derived list. To make up his control condition, he maintained the same first and last items and sampled randomly from the remaining pool of 84 items to construct six 16-item control series. This procedure minimized the effects of remote associations while, at the same time, it served the purpose of equating control and experimental series for item familiarity as well as for the repetition of the initial and final syllables in the original and derived lists. Since Ebbinghaus' control list was not the same as Irion's and since Ebbinghaus' measure of savings was appropriate, the inconsistency between the savings found by Ebbinghaus and the negative transfer found by Hakes et al. (1964) and by Young et al. (1965) remains in need of explanation.

---

[1] This paper was prepared while the first author held a National Science Foundation Traineeship under Grant GZ-1754.

[2] Request for reprints should be sent to Sheldon M. Ebenholtz, Department of Psychology, Psychology Building, Charter at Johnson Streets, Madison, Wisconsin 53706.

[3] In the study Ebbinghaus used total time to learn rather than the number of list repetitions as his dependent measure. In his various experiments, he used both kinds of measures and found they were related enough so that for most purposes it was immaterial which was used. Precautions were taken to increase the accuracy of the time-to-learn measure. For example, he read at a constant rate paced by the ticking of a watch, and he allowed very little delay between study and test. On this latter point, Ebbinghaus (1913) wrote, "There was a perfectly free interchange between the reading and the occasionally necessary tests of the capacity to reproduce by heart. For the latter there was an important rule to the effect that upon hesitation the rest of the series was to be read through to the end before beginning it again [p. 24]."

The theoretical significance of this inconsistency has not yet been determined, but because the inference of negative transfer is clearly required by interference theory, it seems worth considering possible causes of the discrepant results. Among these are the unique features of Ebbinghaus' procedures, some of which have been pointed out by Hakes et al. (1964) and Slamecka (1964): (*a*) Ebbinghaus learned by the method of whole presentation and may have formed some remote associations directly by inadvertently attending to nonadjacent items. This explanation seems quite unlikely in view of the great care that Ebbinghaus took in controlling his reading rate which closely approximated .4 second per syllable. This rapid presentation rate, incidentally, has not been duplicated in contemporary studies of derived lists. (*b*) Ebbinghaus read the list in a regular rhythmical manner. To some extent this rhythmical organization may have contributed to the overall savings, but an examination of the pattern of stressed syllables that he reported revealed no differential effects that would favor lower over higher order derived lists. (*c*) Ebbinghaus served as the sole subject and therefore knew the types of derived lists in the experiment. Slamecka (1964) contended that this is the most likely explanation of Ebbinghaus' results. To support his contention, Slamecka (1964) argued that knowledge of the derived lists is sufficient to produce savings. He found savings for a group who knew the nature of the derived list and no savings for a group that did not. However, besides telling the subjects the nature of the derived list, he also numbered the items successively in the original list and retained the same numbers in the derived list. The presence of these numbers may have allowed subjects to form direct associations between the numerals and the syllables. Knowing the pattern of the numerals, subjects may have provided their own stimuli and transformed the derived list into a paired-associate task. Therefore, Slamecka's (1964) conclusions must be tempered by this fact and also by the fact that Goldstein (1950) told her subjects the nature of a first-order derived list immediately before it was presented and yet the derived list did not show significantly more savings than a scrambled list. Even more importantly, it should be noted that Ebbinghaus himself was aware of this criticism and ran a second experiment in which he prepared the list far enough in advance of running so that he felt successful in eliminating the knowledge of the kind of transformations for any particular list. The savings in this second experiment were only slightly reduced and were comparable to the first. In view of Ebbinghaus' record as a careful experimenter, it seems unlikely that his results can be attributed to this artifact.

There are three additional major differences between Ebbinghaus' experiment and more recent replications which should be considered. First, Ebbinghaus learned derived lists 24 hours after the first list, whereas in recent studies showing negative transfer (e.g., Young et al., 1965), the derived lists were presented a few minutes after the initial lists had been learned. This interlist interval may be a critical factor in accounting for negative transfer obtained in modern studies and should be examined. A second difference is that Ebbinghaus had only occasional tests whereas modern investigators give tests every trial. Izawa (1966) has shown the importance of test trials in learning paired-associate lists. Test trials may also be important in serial learning and perhaps their presence interacts with interference effects. The third difference concerns subject's level of sophistication in dealing with the material to be learned. Ebbinghaus was a highly trained learner whereas most modern experiments have used naive subjects. Ebbinghaus mastered many lists each day and might well have learned to cope with the effects of interlist interference. It may be that extensive practice on A-B, A-D paradigms results in the development of special strategies that reduce interference effects.

Clearly there are many differences between the derived-list study by Ebbinghaus in which he obtained positive transfer and the modern experiments, such as the one by Young et al. (1965) in which negative transfer was obtained. It remains to be determined which of these differences can account for the inconsistent results. It is important to recognize this unresolved problem, especially because Ebbinghaus' findings are at odds with inferences from modern interference theory.

REFERENCES

EBBINGHAUS, H. *Uber das Gedachtnis: Untersuchungen zur experimentellen Psychologie.* Leipsig: Duncker & Humblot, 1885. (Trans. by H. A. Ruger & C. E. Bussenius as *Memory: A contribution to experimental psychology.* New York: Columbia University, Teachers College, Bureau of Publications, 1913.)

GOLDSTEIN, N. Saving in the learning of scrambled lists as influenced by degree of familiarity with nonsense syllables. *Journal of General Psychology,* 1950, **42**, 87–95.

HAKES, D. T., JAMES, C. T., & YOUNG, R. K. A re-examination of the Ebbinghaus derived-list paradigm. *Journal of Experimental Psychology,* 1964, **68**, 508–514.

HAKES, D. T., & YOUNG, R. K. On remote associations and the interpretation of derived-list experiments. *Psychological Review,* 1966, **73**, 248–251.

IRION, A. L. Retroactive inhibition as a function of the relative serial positions of the original and interpolated items. *Journal of Experimental Psychology,* 1946, **36**, 262–270.

IZAWA, C. Reinforcement-test sequence in paired-associate learning. *Psychological reports,* 1966, **18**, 878–919.

POSTMAN, L. Hermann Ebbinghaus. *American Psychologist,* 1968, **23**, 149–157.

SLAMECKA, N. J. An inquiry into the doctrine of remote associations. *Psychological Review,* 1964, **71**, 61–76.

YOUNG, R. K., HAKES, D. T., & HICKS, R. Y. Effects of list length in the Ebinghaus derived-list paradigm. *Journal of Experimental Psychology,* 1965, **70**, 338–341.

(Received March 10, 1971)

Psychologcial Review
1971, Vol. 78, No. 6, 556

LETTERS TO THE EDITOR

INTERDEPENDENCE OF STEVENS' EXPONENTS AND DISCRIMINABILITY MEASURES

Teghtsoonian (1971) presented data which showed that the subjective counterpart to the Weber fraction is a constant independent of sensory modality. He suggested, on the basis of this, that all physical dimensions are mapped into a common continuum of sensory intensity. He also pointed out (see p. 78) that his data are consistent with Fechner's original idea about scaling, namely, that stimuli equally discriminable from matched stimuli will also be matched, the statement of which need not require any assumption that equal just noticeable difference (jnd) units constitute subjectively equal intervals. Fechner's assumption, stated in terms of a cross-modality matching experiment, reads this way: If stimuli I_i' and I_j', drawn from modalities i and j, are equally discriminable from matched stimuli I_i and I_j, also drawn from modalities i and j, then stimuli I_i' and I_j' will be matched also. The relationship between Stevens' exponents and discriminability measures which Teghtsoonian discovered may be derived from this statement in the following formal way:

Suppose I_i' and I_j', equally discriminable from I_i and I_j, are n jnd's from I_i and I_j. It follows that

$$\begin{aligned} I_i' &= (1 + W_i)^n I_i \\ I_j' &= (1 + W_j)^n I_j \end{aligned} \qquad [1]$$

where W_i and W_j are the Weber fractions for modalities i and j. Equation 1 implies that

$$\log(I_i'/I_j)^{(1/n \log(1+W_i))} = \log(I_j'/I_j)^{(1/n \log(1+W_j))}. \qquad [2]$$

Equation 2 enables us to derive the interrelationship between Stevens' law exponents and Weber fractions, for if

$$s(I) = AI^B \qquad [3]$$

where $s(I)$ is the sensation magnitude of a stimulus of intensity I, then I_i' and I_j' will be matched whenever

$$(I_i'/I_i)^{B_i} = (I_j'/I_j)^{B_j}. \qquad [4]$$

Comparing Equation 4 with Equation 2 reveals the interdependence between Stevens' law exponents and Weber fractions, namely, that

$$B_i \log(1 + W_i) = K \qquad [5]$$

where B_i and W_i are the Stevens' law exponent and Weber fraction for modality i, and K is a constant which depends on the reference modality used to determine the Stevens' law exponent.

REFERENCE

TEGHTSOONIAN, R. On the exponents in Stevens' law and the constant in Ekman's law. *Psychological Review*, 1971, **78**, 71-80.

(Received June 17, 1971)

Carl Auerbach
Department of Psychology
Yeshiva University
New York, New York 10003

Abilities: Their Structure, Growth, and Action
Raymond B. Cattell, University of Illinois

Professor Cattell examines the field of abilities from the perspective of personality and motivation research and integrates it with physiological and neurological psychology, comparative animal psychology, machine simulation, and developmental psychology.

583 pages / 1971 / $17.50

Mind in Evolution: An East-West Synthesis of Learned Behavior and Cognition
Gregory Razran, Queens College

Mind in Evolution is an evolutionary synthesis of the main categories of systematic psychology: Russian Pavlovianism, American behaviorism, and Western cognitionism. Professor Razran has unified the once seemingly separate evidence and logic of these major schools into one hierarchical system using the framework of Darwinian evolution.

430 pages / 1971 / $16.50

Personality Development: From Infancy to Adulthood
Charles Wenar, The Ohio State University

Organized by topic rather than age groups, Personality Development is a developmental study of the growth of affection, initiative, self-control, sexual impulses, and other personality variables from infancy to adulthood. Focusing on the individual child and his encounters with family and peers, the author uses a variety of schools of psychological thought as sources of theory and draws on data from case histories, observational research, and controlled studies.

454 pages / 1971 / $8.95

Houghton Mifflin
Publisher of The American Heritage Dictionary of the English Language. Boston 02107 / Atlanta 30324 / Dallas 75235 / Geneva, Ill. 60134 / New York 10036 / Palo Alto 94304

in
the
om

The British Journal of Psychology

Volume 62 Part 3 August 1971

CONTENTS

R. B. JOYNSON Michotte's experimental methods

J. M. CLEGG Verbal transformations on repeated listening to some English consonants

WILLIAM G. CHASE AND HERBERT H. CLARK Semantics in the perception of verticality

E. C. DALRYMPLE-ALFORD Some further observations on the measurement of clustering in free recall

ARTHUR I. SCHULMAN Recognition memory for targets from a scanned word list

H. SCHONE AND N. J. WADE The influence of force magnitude on the perception of body position. II. Effect of body posture

RICHARD F. CROMER The development of the ability to decenter in time

JOHN N. NICHOLSON AND JEFFREY A. GRAY Behavioural contrast and peak shift in children

COLIN D. ELLIOTT Noise tolerance and extraversion in children

P. R. F. CLARKE The Pinmen Recall Test as a measure of the individual

R. P. POWER AND K. D. MACRAE Detectability of items in the Eysenck Personality Inventory

D. E. BLACKMAN AND G. V. THOMAS Critical notice: Reinforcement and information

JOHN ANNETT Reinforcement and information: a reply

PAPERS TO APPEAR IN FORTHCOMING ISSUES

BOOK REVIEWS

OTHER PUBLICATIONS RECEIVED

CHANGE-OVER TO SI UNITS

£2.50 (US $8.50) each, plus postage Annual Subscription £7.50 (US $25.00)

CAMBRIDGE UNIVERSITY PRESS
Bentley House, 200 Euston Road, London NW1 2DB
American Branch: 32 East 57th Street, New York, NY 10022

CONTENTS

PART I

Meditation: Its Spirit and Techniques

CLAUDIO NARANJO

Introduction 3
1. The Domain of Meditation 6
2. Concentrative or Absorptive Meditation 19
3. The Negative Way 75
4. The Way of Surrender and Self-Expression 90
Epilogue 131

PART II

The Techniques of Meditation and Their Implications for Modern Psychology

ROBERT E. ORNSTEIN

Foreword 137
Introduction 139
1. "Turning Off" Awareness 142
2. The Esoteric and Modern Psychologies of Awareness 170
3. An Extended Concept of Human Capacities 213
A Closing Note 233

Notes 235
Bibliography 247

ow, from The Viking Press/
salen Institute Book Publishing Program

N The psychology
of meditation

keystone work in the study of the experimental,
ychological, and physiological aspects of the total
enomenon of meditation by Claudio Naranjo
Robert E. Ornstein An Esalen Book
.95

THE VIKING PRESS

Attc Pocket 2
Issue to Librarian
R Dai 24/3/72

PSYCHOLOGICAL REVIEW

GEORGE MANDLER, EDITOR
JEAN M. MANDLER, ASSOCIATE EDITOR
UNIVERSITY OF CALIFORNIA, SAN DIEGO

Vol. 78, No. 5 **September 1971**

Articles

Organization and Discrepancy in Cognitive Structures **N. T. Feather** 355

The Duplication Theorem of Social Relationships as Tested in the General Population...... **Walter Toman** 380

Components of Attention **Michael I. Posner and Stephen J. Boies** 391

Statistical Decision Model for Auditory Word Recognition........................... **John R. Frederiksen** 409

On the Word Frequency Effect: Comments on the Papers by J. Catlin and L. H. Nakatani **Michel Treisman** 420

Issues in Psychophysical Measurement..**S. S. Stevens** 426

Toward a Theory of Early Infantile Autism **Dewey J. Moore and David A. Shiek** 451

Letters to the Editor...................... 457

Published Bimonthly
by The American Psychological Association, Inc.

Editorial Consultants

TRUETT ALLISON
A. D. BADDELEY
A. T. BECK
JEROME BRUNER
DORWIN CARTWRIGHT
CLINTON DESOTO
BRUCE EARHARD
HOWARD EGETH
HANS FURTH
WENDELL R. GARNER
HARRY GOLLOB
ROBERT GOTTSDANKER
JULIAN HOCHBERG
FRANCIS W. IRWIN
LAWRENCE KARLIN
WILLIAM KESSEN
DAVID H. KRANTZ
R. DUNCAN LUCE
RONALD MELZACK
ALLAN F. MIRSKY
WALTER MISCHEL
JOHN MORTON
RAYMOND NICKERSON
ALAN PAIVIO
CARL PFAFFMAN
BERNARD RIMLAND
ELIOT H. RODNICK
RENATO TAGIURI
ROBERT TEGHTSOONIAN
ANNE TREISMAN
AMOS TVERSKY
WILSE WEBB
JOACHIM WOHLWILL
DAVID ZEAMAN

(A large group of psychologists has kindly agreed to assist in the editing of manuscripts submitted to the *Psychological Review*. In each issue the Editor will list in the *Review* those editorial consultants who have helped him process manuscripts for that issue, so that their efforts may be gratefully acknowledged.)

The *Psychological Review* is devoted to articles of theoretical significance to any area of scientific endeavor in psychology. Except for occasional articles, manuscripts exceeding 7,500 words will be discouraged. Ordinarily, manuscripts which consist primarily of original reports of research or of reviews of the literature should be submitted to other journals. Methodological and statistical articles are not appropriate for the *Psychological Review*. See the editorials in the *Review* (1965, 72, 1–2; 1971, 78, 1–2) for further details.

Beginning with the 1971 issue, Letters to the Editor will be published on an experimental basis. Letters not exceeding 250 words on topics of theoretical interest should be submitted in duplicate.

Manuscripts must be accompanied by an abstract of 100 to 120 words typed on a separate sheet of paper. The abstract should conform to the style of *Psychological Abstracts*. Detailed instructions for preparation of the abstracts appeared in the *American Psychologist* (1961, 16, 833), or may be obtained from the APA Central Office.

All manuscripts must be submitted in triplicate. In the reference list, give journal titles in full; do not abbreviate. Figures should be submitted as glossy prints; duplicate figures may be photographic or pencil-drawn copies. Authors should retain a copy of the manuscript in case there is loss in the mail.

Because of the large number of manuscripts submitted, there is an inevitable publication lag of several months. Authors may avoid this delay if they are prepared to pay the costs of publishing their own articles; the appearance of articles by other contributors is not thereby delayed.

Manuscripts should be addressed to the Editor, George Mandler, Department of Experimental Psychology, South Parks Road, Oxford OX1 3PS, England.

Copyrighted material. The following policies govern reprinting or other forms of reproduction of materials copyrighted by the American Psychological Association: Written permission must be obtained for copying or reprinting any tables, figures, or text of more than 500 words in length. Persons or organizations who obtain permission to reproduce any such copyrighted materials must include on the first page of the reproduced materials the exact copyright notice which appeared on the Association's original publication. There will be a charge of $10.00 per page for such permission except when the requester is the author of the article to be reprinted, or when the article is to be reproduced in a limited number of copies solely for instructional purposes.

HELEN ORR
Managing Editor

LORRAINE BIROS
Technical
Editor

HENRY S. SALMON
Advertising and Business
Manager

ROBERT D. GORMAN
Subscription Manager

PUBLISHED BIMONTHLY BY THE
AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
Prince and Lemon Streets, Lancaster, Pa.
1200 Seventeenth St., N. W., Washington, D. C. 20036

$10.00 volume

Second-class postage paid at Lancaster, Pa.

$2.00 issue

Send all communications, including address changes, to 1200 Seventeenth St., N. W., Washington, D. C. 20036. Address changes must arrive by the tenth of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

PSYCHOLOGICAL
REVIEW

VOL. 78, NO. 5

SEPTEMBER 1971

ORGANIZATION AND DISCREPANCY IN COGNITIVE STRUCTURES

N. T. FEATHER[1]

Flinders University of South Australia

Three structural models are described which have been applied to communication, recall, and attribution of responsibility, respectively. The models are stated in terms of graph theory and are used to represent cognitive structures which may be of two types: abstract and perceived. The abstract structures involved in the three models are assumed to be balanced and are conceived as summaries or abstractions from past experience, internally consistent theories about what ought to be the case. The perceived structures, however, may be balanced or unbalanced and are coordinated to the immediate situation. Discrepancy between an unbalanced perceived structure and the corresponding balanced abstract structure is taken as a condition for cognitive adjustment, behavioral adjustment, or both. Possible ways of expanding the three models are discussed. More general theoretical issues are then considered particularly in regard to the problems of change and organization in abstract structures. Finally, related approaches suggesting the emergence of a general paradigm are briefly noted.

It is rare in psychology to discover paradigms that span seemingly diverse areas. Much of the thrust of model building in recent years has involved a turning away from general principles applicable to a wide range of problems to more specialized miniature systems where the limits of application are clearly defined. This strategy of narrowing the focus is certainly not to be condemned. It is probably necessary to concentrate on well-defined subareas of a field if mathematical models are to achieve any degree of sophistication and if they are to enable precise prediction. But one can sometimes be blinded to the existence of general principles by working within a narrow confine. The gains from intricate formalization may be attended by a premature insulation of the worker from other areas of inquiry which he sees as unrelated to his own. Every so often one needs to take the wide view, for surely one of the goals of science is to search for principles that are sufficiently general that they illuminate and knit together a number of separate areas.

The author has taken the wide view on a previous occasion (Feather, 1959) where models from quite diverse areas of psychology such as mathematical decision theory, social learning, risk taking, performance, and level of aspiration were shown to involve basically similar concepts (expectations and values). This expectancy-value paradigm has since been applied to other areas including further aspects of achievement motivation (Atkinson & Feather, 1966) and information-seeking behavior (Feather, 1967a). It is a way of approach that has

---

[1] Requests for reprints should be sent to N. T. Feather, Department of Psychology, School of Social Sciences, Flinders University of South Australia, Bedford Park, South Australia 5042.

© 1971 by the American Psychological Association, Inc.

wide relevance when the analysis of overt goal-directed behavior is under scrutiny.

The present paper is an attempt to take the wide view again. In this case, however, the general paradigm to emerge is one that concerns cognitive as well as behavioral adjustments. Three models are discussed that have been developed to account for the effects of communication, the process of recall, and the nature of attribution of responsibility for success or failure at a task. On the surface, these three areas appear quite distinct, but it is shown that models involving the same general principles may be applied to them all and that the paradigm that emerges is one that deals with adjustment to discrepancy between structures and the need for a person to achieve a relatively stable view of his world under conditions of varying information input.

ABSTRACT AND PERCEIVED STRUCTURES

The three models to be described have been developed by the author over the past few years. Their development has been guided by a basic conviction that an adequate treatment of cognition must come to grips with the problem of structure. Two types of structure may be distinguished, namely, *abstract structure* and *perceived structure.* The distinction between these two types of structure becomes apparent in subsequent discussion in the content of each of the three models, but in general the abstract structure refers to the residue of experience, to an abstract or summary of past encounters, whereas the perceived structure relates to the immediate situation. It is assumed that abstract cognitive structures develop as the person copes with the influx of information from the external world, that these abstract structures must be organized and relatively stable to provide continuity and meaning under changing circumstances, but that they are susceptible to change as the person matures and as new and discrepant information is received that cannot readily be interpreted in terms of existing abstract structures. These ideas are not new. They are basic to the work of eminent theorists such as Piaget, Bartlett, Tolman, and Bruner. Abstract structures are general schemata that the person carries with him from situation to situation. In contrast, the perceived structure refers to the way information provided by the immediate situation is organized. This type of structure, therefore, is dependent on the immediate situational context and does not have the stable, transituational character of an abstract structure.

The models to be described assume that abstract structures and perceived structures may be defined in terms of the elements they contain and the relations between these elements. All three models involve both personal and impersonal elements; that is, they apply to the person and his relation to information of a personal and/or nonpersonal kind. Two kinds of relations are considered, namely, attitudinal and unit. In the present context, the attitudinal relation refers to the process of *evaluation* and relates any two elements of a structure, one of which must refer to a personal entity. A positive attitudinal relation is taken to imply positive evaluation such as liking, favoring, agreeing with, and a negative attitudinal relation is taken to imply negative evaluation such as disliking, opposing, disagreeing. In the present context, the unit relation refers to *association* between two elements of a structure. These elements may refer to either personal or impersonal entities or both. A positive unit relation is taken to imply positive association or linking together of elements, a negative unit relation is taken to imply dissociation—not just a neutral lack of association, but an active pushing apart or opposition to unit formation between elements. The negative unit relation, therefore, is not the same as a null relation which would imply simple lack or absence of association. The negative unit relation carries with it the definite implication of a resistance to association. These two types of relation (attitudinal and unit) are derived from Heider (1958). The attitudinal relation corresponds to Heider's sentiment relation, involving feeling and evaluation. In the Heiderian analysis, the positive unit relation has been applied to dyads where entities are related through similarity, causality, ownership, or other unit-forming characteristics, but Heider has little to say about negative unit relations. Indeed he does not distinguish clearly between a negative unit relation and the absence of a relation.

The three models to be described therefore involve abstract structures and perceived structures defined in terms of unit and/or attitudinal relations between elements. It is evident that the processes of association and evaluation are held to be basic in the formation of these structures. The person is seen as an active processor of information, associating separate units of information on the basis of similarity, causality, contiguity, ownership, etc., and evaluating information along a good–bad dimension. It is further assumed in each of the three models that the relatively stable abstract structure tends to be organized according to Heider's *principle of balance* but that the perceived structure may be balanced or unbalanced depending on the information given. Although Heider's balance principle is generally familiar, it will be explained briefly in the following section since it is basic to the understanding of the general theory involved in the three models to be discussed.

STRUCTURAL BALANCE

Heider defines his principle of balance in terms of a tendency for unit (U or not U) relations and sentiment (L or DL) relations to fit together harmoniously. The conditions for balance are defined as follows:

> A dyad is balanced if the relations between the two entities are all positive (L and U) or all negative (DL and not U). Disharmony results when relations of different sign character exist. A triad is balanced when all three of the relations are positive or when two of the relations are negative and one is positive. Imbalance occurs when two of the relations are positive and one is negative. The case of three negative relations is somewhat ambiguous . . . [Heider, 1958, pp. 202–203].

If one considers as entities person (P), other person (O), and impersonal object (X), the following are examples of balanced states: P likes something he made ($_{P}L_{X}$, $_{P}U_{X}$) and P likes what his friend likes ($_{P}L_{O}$, $_{O}L_{X}$, $_{P}L_{X}$). Some examples of unbalanced states are as follows: P's friend dislikes him ($_{P}L_{O}$, $_{O}DL_{P}$) and P likes what his friend dislikes ($_{P}L_{X}$, $_{P}L_{O}$, $_{O}DL_{X}$). Heider assumes that states of balance are preferred over disharmony, that new relations may be induced by existing relations so as to produce a balanced state, that where relations are not in balance there will be a tendency for them to change so that a balanced configuration is achieved, and that if such a change is not possible, the state of imbalance will produce tension. Thus the Heiderian view is of structures containing relations that are dynamically interdependent, moving toward harmonious organization.

DISCREPANCY BETWEEN ABSTRACT STRUCTURE AND PERCEIVED STRUCTURE

In the three models to be discussed, it is assumed that the abstract structure based on previous experience is balanced. The perceived structure may be balanced or unbalanced according to the information presented. When both the perceived structure and the corresponding abstract structure are balanced there is no pressure to change, that is, no cognitive adjustment is necessary. But when the perceived structure is unbalanced, then a *structural discrepancy* exists between the perceived structure and the underlying abstract structure and there will be pressure to resolve the discrepancy. This resolution may involve a change in the perceived structure in the direction of the corresponding abstract structure. The present paper focuses on such cognitive adjustments. In other cases, however, direct behavioral adjustments may occur in an attempt either to obtain more information or to escape from the discrepant situation. And in still other cases, where the structural discrepancy persists, changes in the underlying abstract structure may occur. It is assumed that the resolution of structural discrepancy between perceived and abstract structures is necessary if the individual is to achieve a relatively stable view of his physical and social world. It is also assumed that the perceived structure relating to the immediate situation is more amenable to change than the abstract structure which is grounded in the individual's past experience, though the abstract structure may itself change as new and discrepant information is continually experienced. Finally, since it is probably easier to distort social reality than to distort physical reality, changes in perceived and abstract structures keyed to social reality may be more likely to occur than changes in these structures when they are keyed to physical reality.

GRAPH THEORY

In the models to be described, abstract structures and perceived structures are represented in terms of the mathematics of linear graphs (Cartwright & Harary, 1956; Harary, Norman, & Cartwright, 1965). One can map elements and relations into points and lines of a linear graph and define the overall balance within the system. Figure 1 presents a number of structures, half of which are balanced and half of which are unbalanced according to Heider's principle. In these structures, a postive attitudinal relation is represented by a solid line and a negative attitudinal relation by a dashed line; a positive unit relation is represented by a solid bracket and a negative unit relation by a dashed bracket. An arrow on a line indicates the direction of an attitudinal relation. Thus Figure 1 (a) represents a situation in which person (P) likes other person (O), and both like impersonal object (X). This is the familiar balanced triad in which a person likes what his friend likes.

The structures in Figure 1 are called *signed digraphs* (s digraphs), since lines in a structure may have both sign (positive or negative) and direction. In Figure 1, Structures a, b, c, d, e, and f are balanced, and Structures g, h, i, j, k, and l are unbalanced. Also in Figure 1, s digraphs a, c, f, g, h, and l involve only the one type of relation, in each case an attitudinal relation. They are called s digraphs of *Type 1.* The remaining s digraphs involve both attitudinal and unit relations, that is, two types of relation, and are called s digraphs of *Type 2.* Each of the s digraphs in Figure 1 contains a semicycle which is defined by Cartwright and Harary (1956) as "a collection of lines obtained by taking exactly one from each pair $\overrightarrow{AB}$ or $\overrightarrow{BA}$, $\overrightarrow{BC}$ or $\overrightarrow{CB}$, . . . , $\overrightarrow{DE}$ or $\overrightarrow{ED}$, and $\overrightarrow{EA}$ or $\overrightarrow{AE}$ [p. 283]." Each semicycle has a sign which is defined as the product of the signs of its lines. For example, the semicycle $\overrightarrow{PO}$, $\overrightarrow{OX}$, $\overleftarrow{XP}$ is positive in Figure 1 (a) but negative in Figure 1 (g). Cartwright and Harary have generalized Heider's principle by defining an s digraph as completely bal-

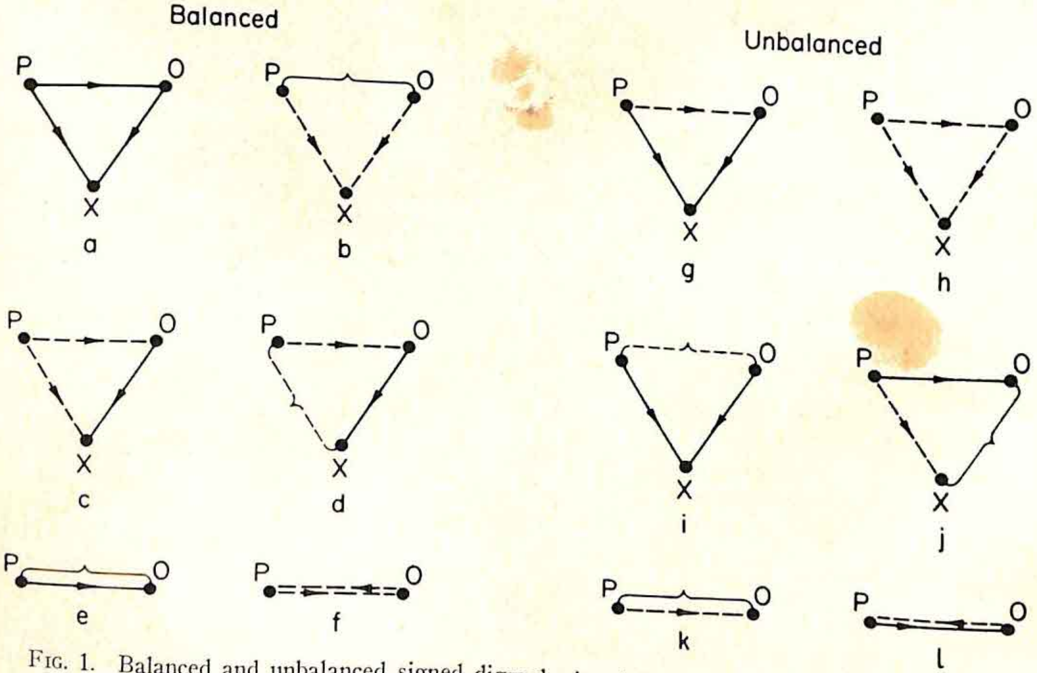


FIG. 1. Balanced and unbalanced signed digraphs involving attitudinal and unit relations.

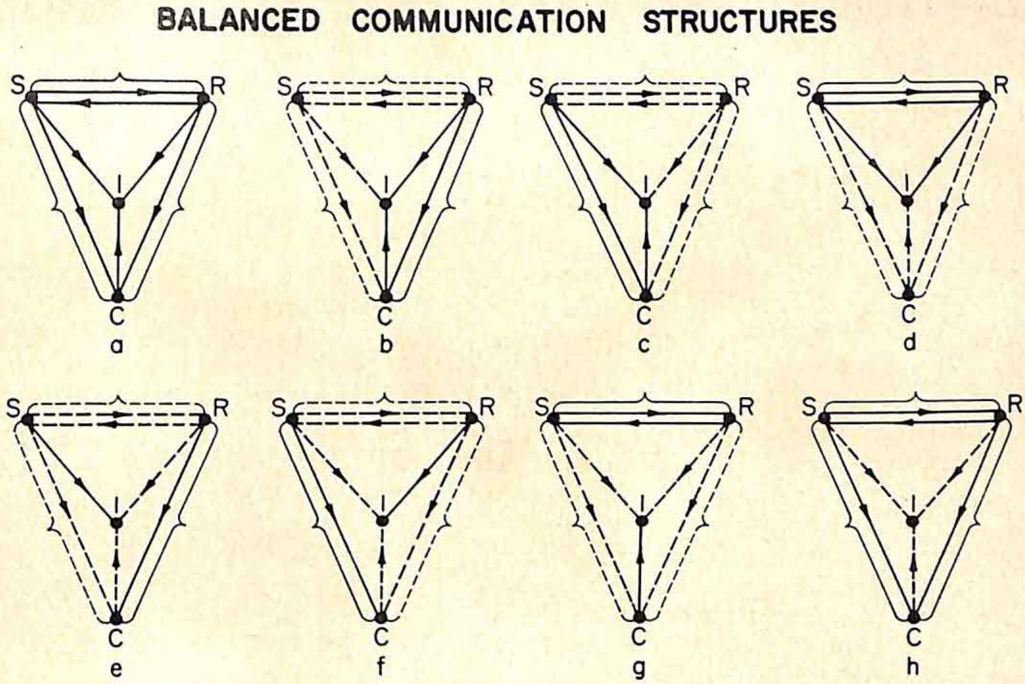


FIG. 2. Signed digraphs for the communication model representing eight completely balanced structures.

anced when the signs of all the semicycles it contains are positive (see also Harary et al., 1965). It can be observed in Figure 1 that each of the balanced s digraphs involves one positive semicycle and each of the unbalanced s digraphs involves one negative semicycle. In more complex structures, there could be a number of semicycles all of which must be positive if the entire structure is balanced.

The three models will now be presented. Since each has been described in detail before, along with the research that it has generated (Feather, 1964, 1967b, 1969a, 1969b), the present statement will be restricted to a short description of each model followed by a general comparison of the models.

THE THREE MODELS

The Structural Balance Model of Communication Effects

The structural balance model of communication effects is presented in Figure 2. Each s digraph in Figure 2 is coordinated to a balanced structure in which the elements refer to a source or communicator (S) who is presenting a communication (C) about an issue (I) to a receiver or communicatee (R). Relations within the structures are unit and attitudinal and can be positive or negative. The elements of the structure are coordinated to points of the s digraph and the unit and attitudinal relations are represented by using positive and negative brackets and lines as described in the previous section. The arrows on the lines describe the direction of each attitudinal relation. It will be observed that all semicycles are positive, that is, in balance.

The s digraphs in Figure 2 may be coordinated both to balanced abstract structures and balanced perceived structures. An example of a balanced perceived structure is one corresponding to an objective situation in which a person presents a communication about some issue to his friend, and both his attitude and the friend's attitude toward the issue are in the same direction as the stand taken in the communication. Under these circumstances there will be no pressure to change, since the perceived structure corresponds to an underlying balanced abstract structure (either s digraph a or s digraph h in Figure 2). An example of an unbalanced perceived structure is one corresponding to an objective situation in which a person presents a communication about some issue to his friend and both his attitude and his friend's attitude toward the issue are in the opposite direction to the stand taken in the communication. Two such unbalanced perceived structures are presented in Figure 3. Under these conditions, a discrepancy exists between the perceived structure (either s digraph a or s digraph b in Figure 3) and the corresponding balanced abstract structure (either s digraph a or s digraph h in Figure 2). There will be pressure to resolve the discrepancy and this may involve some form of cognitive adjustment (e.g., the source's message might be misinterpreted) or overt response (e.g., the source might be asked to clarify his position) or both.

It is important to note that in any communication situation the perceived structure of the source of the communication may be quite different from the perceived structure of the receiver of the communication. Typically the source has more information about why he is performing the act. In situations where he is involved in attitude-discrepant communication, for example, the source may know that he has been coerced in some way to deliver the message, whereas the receiver may be quite ignorant of this fact. In this case the receiver may perceive the source as acting of his own volition and responsible for his act, whereas the source would know better—he may from the start deny responsibility for his attitude-discrepant action.

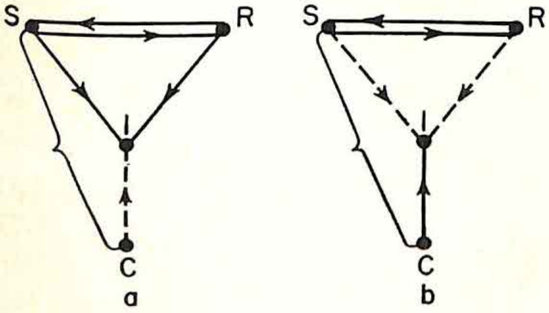


Fig. 3. Signed digraphs representing two unbalanced perceived structures corresponding to communication situations where the communication is discrepant with attitudes of both source and receiver who are mutual friends.

A proper test of the communication model depends on how successfully one can coordinate the theoretical concepts to empirical operations. Positive or negative attitudes between source and receiver, positive or negative attitudes of source or receiver toward the issue, positive or negative stands taken by the communication toward the issue, agreement or disagreement with the communication by source or receiver have all been taken as instances of attitudinal relations. Unit relations have been somewhat harder to specify, particularly when they are assumed to be negative. Membership of a source and receiver in the same tightly knit group has been represented by a positive unit relation linking S and R. A positive unit relation between source and communication has been taken to represent the situation in which the source sees himself as responsible for the communication (source's perceived structure) or the receiver sees the source as responsible for the communication (receiver's perceived structure). The negative unit relation between source and communication has been taken to represent the case where the source sees himself as disowning the communication (source's perceived structure) or is perceived by the receiver as not responsible for the communication (receiver's perceived structure).

The model presented in Figure 2 has considerable predictive power. Provided one knows any *three* relations that involve the *four* elements (S, C, I, R) of a structure, one can use the principle of structural balance to derive the remaining relations. This type of strategy can be employed in studies designed to *map into* the assumed balanced abstract structures. For example, one can present a hypothetical communication situation in which the source's communication favors some issue (positive $\overrightarrow{CI}$ attitudinal relation), the receiver likes the source (positive $\overrightarrow{RS}$ attitudinal relation), and the receiver sees the source as responsible for the communication (positive SC unit relation). One can ask subjects to predict whether the receiver's attitude toward the issue will be favorable (positive $\overrightarrow{RI}$ attitudinal rela-

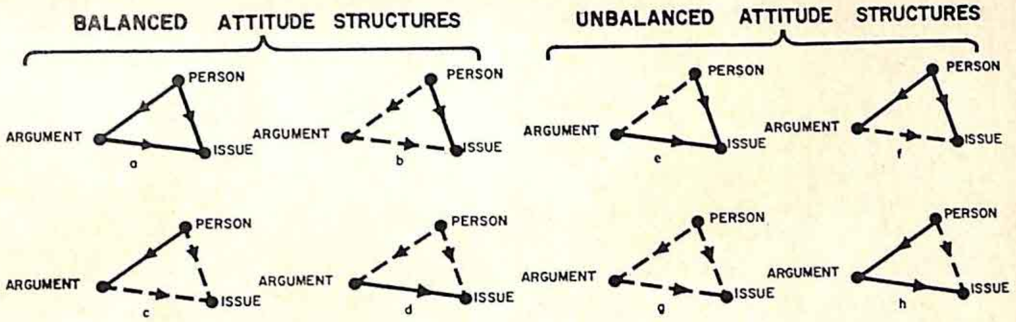


FIG. 4. Signed digraphs for the recall model representing four balanced structures and four unbalanced structures.

tion), whether the receiver will agree with the communication (positive $\overrightarrow{RC}$ attitudinal relation), etc. Such a study would seek to discover if subjects were using the balanced abstract structure represented in Figure 2a in making their judgments. In other investigations, one can design situations in which the perceived structures are unbalanced from the start and study the cognitive adjustments that occur. In fact the structural balance model of communication effects has initiated an extensive program of research (Feather, 1965a, 1965b, 1966, 1967c; Feather & Armstrong, 1967; Feather & Jeffries, 1967). It has also been applied to the literature on attitude change and has proved to be a powerful and general model in its ability to illuminate this highly complex area (see Feather, 1964, 1967b).

Structural Balance Model of Recall

The structural balance model of recall is presented in Figure 4. Each s digraph in Figure 4 is coordinated to a structure in which the elements are *person, argument,* and *issue*. The basic structure is therefore similar to the S-I-C and R-I-C triads in the communication model of Figure 2. The recall model involves only one person, however, and a simple triadic structure that includes positive and negative attitudinal relations. These relations, which are represented by directed positive and negative lines in each s digraph, refer to the person's attitude toward the issue, the person's attitude toward the argument, and the stand taken by the argument toward the issue. For example, Figure 4c could represent an abstract structure in which a person opposes capital punishment (the issue), knows an argument that opposes capital punishment, and agrees with that argument. In Figure 4, half of the s digraphs are balanced; the other half are unbalanced.

The model has been applied to the situation where a person is asked to recall arguments about some controversial issue. Meaningful material, therefore, is elicited toward which the person has a definite attitude. It is assumed that he will employ balanced abstract structures corresponding to s digraphs a, b, c, and d in Figure 4 in recalling arguments, since the structures corresponding to s digraphs e, f, g, and h are unbalanced and unlikely to survive in memory. This assumption is consistent with the general proposition developed previously, namely, that the products of long-term information processing are likely to be abstract structures that are organized rather than disorganized. Hence arguments that are keyed into balanced abstract structures will be more available for recall.

This emphasis on abstract structures as directive influences on the recall process echoes Bartlett's (1932) view:

> when a subject is being asked to remember, very often the first thing that emerges is something of the nature of an attitude. The recall is then a construction, made largely on the basis of this attitude, and its general effect is that of a justification of the attitude [p. 207].

In the present model, the constructive nature of recall is also emphasized (see Feather, 1970; Neisser, 1967, p. 292), and the constructive activity is assumed to be guided by balanced abstract structures in accordance with the demands of the immediate social situation.

The recall model has served as the basis for an extensive program of research (Feather, 1969a, 1969c, 1969d, 1970) mainly concerned with mapping abstract structures in relation to what is recalled. These studies have shown that subjects who were asked to report arguments that were pro and con on an issue (e.g., Vietnam) presented arguments that tended to be consistent with their own attitude toward the issue, where consistency was defined as in Figure 4. Thus a person in favor of an issue would report a large number of pro arguments with which he agreed and a large number of con arguments with which he disagreed (see balanced s digraphs a and b in Figure 4), but very few pro arguments with which he disagreed or con arguments with which he agreed (see unbalanced s digraphs e and f in Figure 4). Thus his recall was selective and appeared to be determined by balanced abstract structures.

The degree to which cognitive structures facilitate recall has been investigated in other contexts apart from the present one which involves controversial issues toward which the person has a definite attitude and for which the special recall model illustrated in Figure 4 was formulated. For example, Deese (1965) has investigated associative structures using the method of word association, and Bousfield (1953) has examined the extent to which words are recalled in meaningful clusters (for a recent review of clustering in recall see Shuell, 1971). Each of these different approaches emphasizes the concept of underlying structures directing the recall process. The degree to which these underlying structures would involve attitudes would vary according to whether the material was personally relevant (as as-

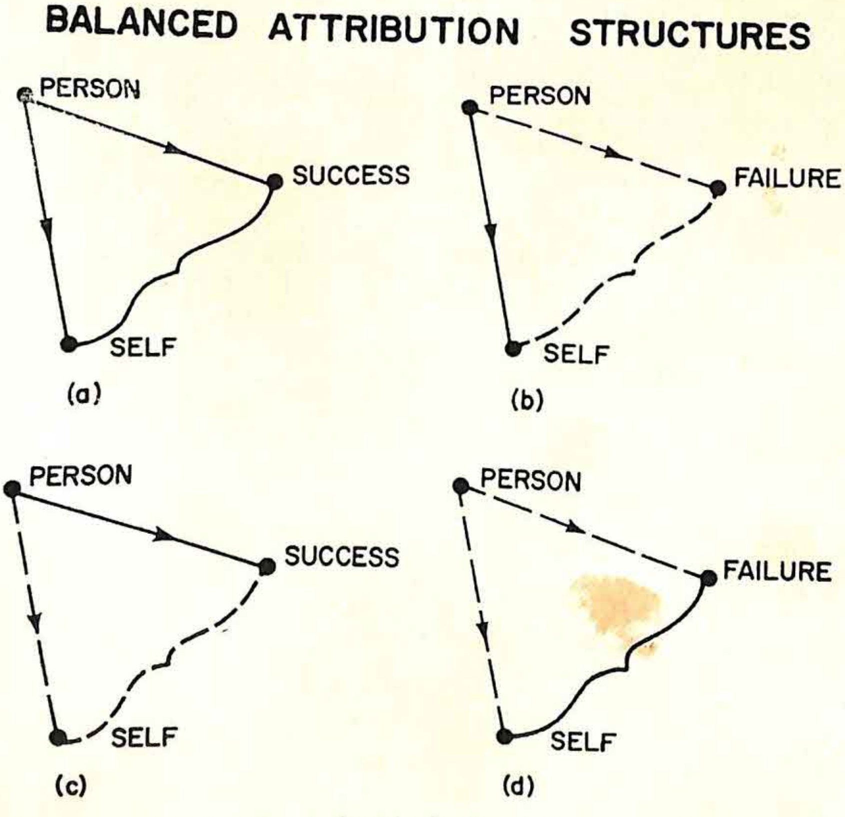


FIG. 5. Signed digraphs for the attribution model representing four balanced structures.

sumed in the present model) or relatively neutral (as assumed in most of the studies of "clustering").

Structural Balance Model of Attribution of Responsibility

The structural balance model of attribution of responsibility is presented in Figure 5. Each s digraph is coordinated to a structure in which *person, self,* and *outcome* (success or failure) are the elements and the relations are unit and attitudinal. As before, the elements of the structure are mapped into points of the s digraph, the attitudinal relations into directed lines which may be positive or negative in sign corresponding to the direction of the attitude (pro or con), and the unit relations into positive or negative brackets. The model assumes that success is positively evaluated by the person, whereas failure is negatively evaluated. But the person may have either a favorable or an unfavorable attitude toward himself. The positive unit relation between self and outcome has been taken to imply that the self is seen as owning the outcome—responsibility for the outcome is attributed by the person to internal factors within the self such as ability. The negative unit relation between self and outcome has been taken to imply that the self is seen as disowning or not responsible for the outcome—responsibility for the outcome is attributed to factors outside the self such as luck. This coordination of perceived responsibility to the positive unit relation and perceived lack of responsibility to the negative unit relation is consistent with the interpretation of the positive or negative unit relation between source and communication discussed in relation to the communication model. The attribution model, however, involves only one person, not two as in the communication model.

The four s digraphs in Figure 5 may be coordinated both to balanced abstract structures and balanced perceived structures. An example of a balanced perceived structure would be one coordinated to a situation in which a person with strong negative self-evaluation fails at a task and sees his failure as due to lack of ability. Under such conditions there would be no pressure to change the perceived structure, since it corresponds to an underlying balanced abstract structure (s digraph d in Figure 5). An example of an unbalanced perceived structure would be one in which a person with high positive self-evaluation succeeds at a task but is led to believe that his success was due to external factors such as good luck and not to his own ability (see Figure 6). In this case there will be a discrepancy between the perceived structure represented in Figure 6 and a corresponding abstract structure (s digraph a or s digraph c in Figure 5). There will be pressure to resolve the discrepancy and this may involve some form of cognitive adjustment. For example, the person might decrease his self-evaluation (bringing about correspondence with s digraph c) or restructure the situation so that the success can be attributed to self after all (achieving correspondence with s digraph a).

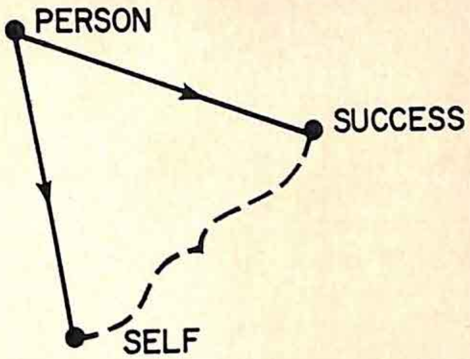


Fig. 6. Signed digraph representing an unbalanced perceived structure corresponding to a situation where success is attributed to external factors.

The attribution model has also led to an extensive program of research (Feather, 1969b; Feather & Simon, 1971a, 1971b), most of it involving mapping into the four balanced abstract structures assumed in the model. One of the main predictions so far tested has been that a person will attribute good outcomes (success) to himself when self-evaluation is positive but will disown good outcomes when self-evaluation is negative. In contrast, bad outcomes (failure) will be attributed to the self when there is negative self-evaluation but will be disowned when there is positive self-evaluation. This hypothesis has been supported in a specific task situation in which high expectations of success have been taken to connote positive self-evaluation and low expectations of success to connote negative self-evaluation with respect to the task to be performed. Thus a highly confident person (positive self-evaluation) tended to attribute the outcome more to ability than to good luck if he succeeded, and more to bad luck than to lack of ability if he failed. An unconfident person (negative self-evaluation) tended to attribute the outcome more to good luck than to ability if he succeeded, and more to lack of ability than to bad luck if he failed. Attribution of responsibility for success and failure, therefore, appeared to follow balanced abstract structures in terms of the s digraphs in Figure 5.

Summary

The three models that have been described are the products of an overall strategy that attempts to deal with perceived and abstract structures by specifying elements and relations and by invoking the principle of balance to account for the ways in which the abstract structures become organized. Unbalanced perceived structures are assumed to be under pressure to change toward corresponding balanced abstract structures and any discrepancy may result in cognitive or behavioral adjustments on the part of the person aimed at reducing the discrepancy.

The three models involve unit and attitudinal relations and thereby recognize the importance of the processes of association and evaluation in the formation of structures. But the attitudinal relations have been included only because each model involves personal elements. One can also conceive of structures, where evaluation is not important and where the elements are impersonal, associated in terms of unit-forming characteristics such as similarity and proximity. One's cognitive map of the physical structure of a town or city would be an example.

STRUCTURAL CORRESPONDENCE AND STRUCTURAL DISCREPANCY

In presenting the three models described in the previous section, it was implicitly assumed that the elements in the perceived structure would correspond to the elements in the underlying abstract structure; for example, in the case of the communication model, that both perceived structure and abstract structure would contain the same set of elements, namely S (source or communicator), C (message or communication), I (issue), and R (receiver or communicatee). In the sense that the abstract structure is considered to be an abstraction or summary of previous experience, this assumption appears to be reasonable. One can conceive of a person forming a simplified schema in which the common elements of a situation are abstracted over time and become incorporated into an abstract structure. It would be difficult, for example, to imagine a communication situation that did not involve the four elements mentioned above. In terms of the logic of abstraction, therefore, one would expect correspondence or isomorphism betweeen elements of the abstract structure and elements within the perceived structure. Indeed, over time the abstract structure may come to exert a directive influence on the way in which the actual situation is perceived, permitting a simplified view of present reality in terms of the basic elements involved in the abstract structure. Also over time the abstract structure may be enlarged as more common elements are abstracted on the basis of experience.

What then is the basis of *structural discrepancy* if there is correspondence between elements of the abstract structure and elements within the perceived structure? Clearly, structural discrepancy would depend on the relations rather than the elements within the perceived structure, and increasing discrepancy would occur as more and more relations in the perceived structure had to be changed in order to correspond to the underlying abstract structure. In these terms, one index of the size of structural discrepancy would be the minimum number of relations that would have to be modified if the perceived structure were to correspond to an underlying abstract structure. Such an approach would have the effect of equating size of structural discrepancy with the degree of balance in the perceived structure (see Harary et al., 1965, Ch. 13), size of discrepancy increasing as degree of balance becomes less. In Figure 7, for example, two relations would have to be altered to bring s digraph b into correspondence with one of the three possible abstract structures. Structural discrepancy would be greater in this case than for s digraph a where only one relation would have to be changed to accommodate to the corresponding abstract structure. Signed digraph a could represent the situation where source (S) and receiver (R) are mutual friends with similar attitudes toward an issue (I), and where source (S) delivers a message or communication (C) that is discrepant with his own attitude toward the issue. The situation represented by s digraph b would be similar except that in this case, the two people are not mutual friends, S likes R, but R does not like S. Intuitively the latter situation appears more discrepant from a balanced state than the former, and indeed, in terms of the above definition, it is more structurally discrepant.

The above definition of structural dis- crepancy, however, is an oversimplification for it neglects the possibility that relation may differ in strength as well as in sign. A single relation may be very difficult to chang because it is strong (e.g., intense liking fo a friend), whereas two relations may b quite easy to change because they are weak Hence an index of structural discrepanc should probably refer not only to the min mum number of relations that have to b altered to produce balance, but also to th *strength* of those relations relative to othe relations in the perceived structure (i.e., t number weighted by strength).

Finally, there may be a tendency for ne relations to be *induced* in the perceived stru ture in line with the underlying abstra structure, that is, the abstract structure ma determine the induction of new perceive relations involving association or evaluatio or both in their various forms. For e ample, a person listening to a friend's con munication that takes a position consona with his own attitude may perceive his frien as having the same attitude toward the iss as he has himself.

The questions raised in the present sectio suggest some new directions in the analy

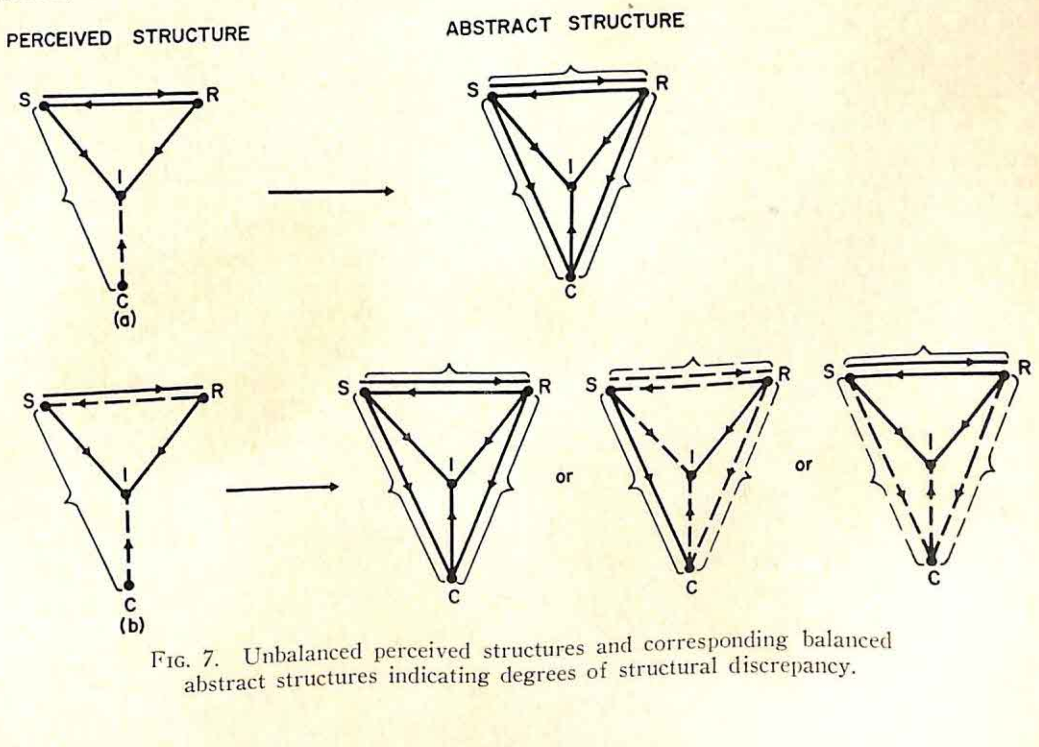


FIG. 7. Unbalanced perceived structures and corresponding balanced abstract structures indicating degrees of structural discrepancy.

of cognitive structures, particularly in regard to the specification of structural discrepancy. In the next section various way in which the three models may be further expanded are considered and problems that require future solution are briefly discussed.

FURTHER EXPANSION OF THE THREE MODELS

Elements and Relations

The three models so far explored have been extremely simple ones involving no more than four elements and no more than two relations. In actuality, however, situations are likely to be much more complex so that, in a communication situation, for example, a receiver may be exposed to a number of communications about a variety of issues and he may be only one of a large audience listening to the source. It follows that perceived structures may be quite complex, involving a constellation of elements and relations. Abstract structures, on the other hand, are likely to be simpler in their formal properties since they are the residues of experience, summaries or abstractions from what has gone before.

It should be possible to expand the models described previously by adding new elements and new relations. One such extension has already been made in an attempt to conceptualize the concept of "importance of the issue" in the communication model. The concept of importance is one that has been introduced into a variety of theories. For example, Festinger (1957) allowed for such a concept when he included the importance of the cognitive elements involved in a dissonant relationship as a factor influencing the magnitude of dissonance. In relation to the communication model (Feather, 1967b), it has been argued that one can represent an important issue "as one toward which the person has a strong attitude and as one that is closely bound together with other issues toward which the person also has strong attitudes of the same sign [p. 149]." Figure 8 represents such a situation in which Issues I, J, and K are important to receiver R, and in which a communication C about issue I is sent by a positive source S. The stand taken in this communication is discrepant with R's attitude toward I in both s digraph a and s digraph b. Obviously the perceived structures corresponding to s digraph a and s digraph b are unbalanced. It would be possible to restore balance in the S-C-I-R semicycle if the $\overrightarrow{RI}$ attitudinal relation were to change from positive to negative (s digraph a) or from negative to positive (s digraph b). One would expect such changes to be resisted, however, since they would throw the semicycles involving I, J, and K out of balance. Thus, in the two cases represented, the receiver might dissociate the source from the communication (a rather difficult cognitive adjustment to make, see Feather, 1965b), he might misperceive the stand taken by the communication (if there is some ambiguity) or he might change his attitude toward the source. All of these would be effective cognitive adjustments that would resolve the imbalance.

It will be noted that the above interpretation of importance of the issue preserves the structural character of the communication model and assumes that balanced abstract structures exist that are somewhat more complex than those represented in Figure 2. Other possible expansions of the communication model might include additional elements to allow for the group memberships of the source and of the receiver, and additional elements to represent the value base of the source and the receiver (perhaps enabling better representation of the perceived trustworthiness of the source in terms of the values he holds). Additional relations might also be incorporated to take account of dominance or power relationships between source and receiver (e.g., the authority of the source) and the role characteristics of source and receiver (perhaps by integrating the communication model with structural role theory; Oeser & Harary, 1962, 1964). Similarly, in the recall model, one might include extra elements and unit relations to represent clusters of issues and arguments, and in the attribution model, additional elements to represent other persons when attribution is investigated within a social context (Feather & Simon, 1971a). Such extensions are feasible and would enable ex-

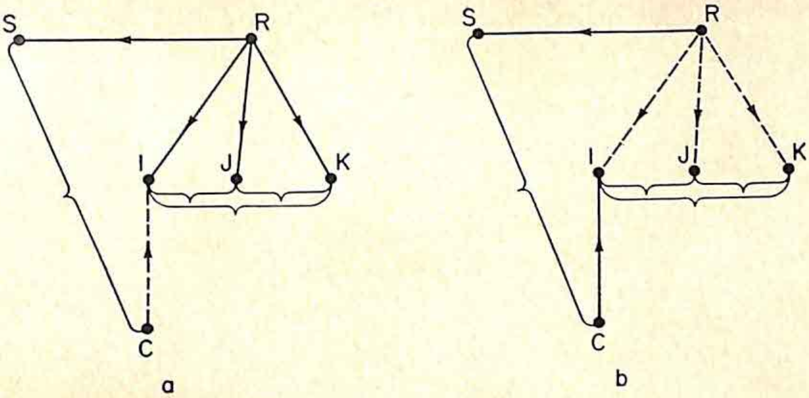


FIG. 8. Unbalanced signed digraphs in which Issues I, J, and K are important to receiver R.

pansion of the structural models to wider areas while still operating within the mathematical framework of linear graph theory.

It is apparent that one psychological factor limiting the size and complexity of abstract structures in the three models presented is the general assumption of balance in the underlying abstract structures. It is probably easier to justify this assumption in those cases where the number of elements and relations and hence the number of semicycles is small (as in the three models presented). Large, complex abstract structures may not be completely balanced, however, since they would contain large numbers of semicycles, some of which would involve many elements, and all of these semicycles would have to be balanced—a formidable task for the individual. For this reason, abstract structures of the kind presented in the three models may involve only a limited number of elements. Highly complex structures are probably beyond the capacity of most individuals, and simplification of complex information input so that it becomes coded in terms of smaller organized units is probably the normal course of events. As has been argued, this simplification is an important part of adaptation, enabling the individual to achieve stability in the face of complexity and variation.

However, it would be foolish to ignore the fact that individuals do differ in their capacity to cope with the welter of information they receive and that some seem able to develop more complex and differentiated abstract structures than others. Since it would be difficult to balance these structures completely, other types of balance such as *local balance* may become important (see Harary et. al., 1965). Local balance would occur, for example, if only the semicycles involving the *self* as an element were balanced—a plausible assumption in more complex structures where *self* is an element. Whether there are general principles that underly the particular inconsistencies that can be tolerated in complex structures is an important question. The articulation of such principles as they might apply to structures of greater complexity is a task for the future, demanding both theoretical and empirical inquiry. Psychologically, however, it is important to allow in any conceptualization for people who have a complex view of existence that tolerates some inconsistencies. And, in line with the above discussion, some inconsistencies may be almost inevitable when cognitive structures are complex.

Strength of Relations

Expansion of the present models requires some means of representing the *strength* of different relations, since at present the models only enable predictions about *sign* (positive or negative). Balance theory is in this respect somewhat akin to a two-valued logic when what is required in expanded models is a multivalued logic. The present approach enables one to deal with sets of relations such as: If P likes O and if P likes X then O will like X, but it does not enable one to answer questions such as: If P likes O with probability p_1 and P likes X with probability p_2, what is the probability that O will like X? The concept of strength will be necessary if one is to predict the strength of new induced relations in a perceived structure and if one is to arrive at an adequate index of structural discrepancy. More generally, the question of whether there are equilibrium states for strengths of structural relations as well as for signs of structural relations (as in balanced structures) is an intriguing one, deserving of close investigation.

In regard to the *induction* of new relations, several principles have been advanced by the author in a study of how subjects predict interpersonal attraction (Feather, 1966). At the level of the dyad, a *similarity principle* appeared to hold. Thus, in two person structures involving person (P) and other person (O), a predicted attitudinal relation tended to mirror a given unit or attitudinal relation in strength as well as in sign. For example, if a subject knew that P liked O a lot or that P and O both belonged to the same tightly knit group, there was a tendency for him to predict that O will very probably like P. At the level of the triad, a *discrepancy principle* taking account of both sign and strength of relations seemed to fit the data best. This principle asserts that when two given relations have the same sign, large discrepancies in their strength will be associated with a weak positive third relation and small discrepancies with a strong positive third relation completing the triad. When two given relations are of *opposite* sign, large discrepancies in their strength will be associated with a strong negative third relation and small discrepancies with a weak negative third relation.

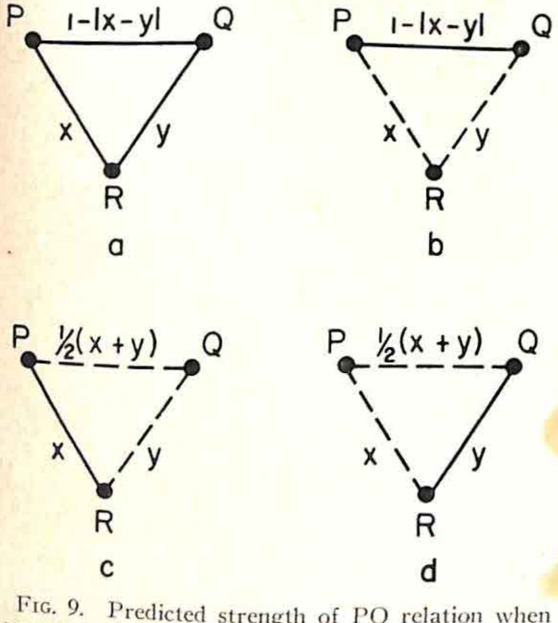


FIG. 9. Predicted strength of PQ relation when strengths of PR and QR relations are given using the discrepancy principle for given relations of same sign (signed digraphs a and b) and unlike sign (signed digraphs c and d).

The discrepancy principle may be applied to all triads in the three models discussed previously and is generalized mathematically in Figure 9 which represents any triad of relations containing elements P, Q, R. It is assumed in Figure 9 that relations PR and QR are given and that PQ has to be predicted. As before, solid lines represent positive relations and dashed lines represent negative relations. The numbers x and y refer to the strength of the PR and QR relations, respectively, and are assumed to fall between 0 and 1. The predicted strength of the PQ relation according to the discrepancy principle is indicated in Figure 9 for triads containing given relations of the same sign (s digraphs a and b) and of opposite sign (s digraphs c and d). The predicted strength will be between 0 and 1.

The two principles mentioned above are extremely simple ones and will undoubtedly require modification as they are tested empirically. Even now, it will be apparent that the discrepancy principle leads to paradoxical predictions as x and y both approach zero, since the third relation would approach 1 when the two given relations are of the same sign and 0 when the two given relations are of opposite sign (see Figure 9). Moreover, one needs a formulation of strength that can be applied in a mathematically consistent way to structures containing any number of elements and relations. The problem is complicated by summation effects (Feather, 1966) that may occur in structures. The possibility exists that there may be no general mathematical solution to these problems, but this has still to be shown. If there is no general solution, one will have to be content with structural models that predict sign but not strength of relations, an important achievement nevertheless.[2]

Other Forms of Cognitive Bias

While research based on the three models discussed in the present paper has strongly supported the assumption of underlying balanced abstract structures, the results have also shown that it may be necessary to consider the effects of other forms of cognitive bias. Chief among these has been a positivity bias: a tendency for subjects to bias evaluation in the positive direction. In tests of the three models, this effect has appeared in a tendency (*a*) for the receiver of a communication to evaluate the source of the communication positively (Feather & Armstrong, 1967; Feather & Jeffries, 1967), (*b*) for persons to recall more arguments they agreed with than arguments they disagreed with (Feather, 1970), and (*c*) for a person to attribute the success of another person more to ability than one's own success, and the failure of another person less to lack of ability than one's own failure (Feather & Simon, 1971a; Simon, 1969). Positivity effects have also been found in other areas of social psychology (Feather, 1967b; Zajonc, 1968). Indeed Rosenberg (1968) has recently referred to "the individual's need to so construct 'reality' as to emphasize hopeful prospects [p. 91]," a psychological need toward wish fulfillment that may interfere with the arousal of inconsistency and the restoration of consistency. Earlier, Rosenberg and Abelson (1960) had spoken of a tendency "to maximize expected gain and to minimize expected loss [p. 145]." In addition, other forms of cognitive bias have been explored (e.g., reciprocity, agreement, extremity). Some of these effects may be situation specific, but others (like positivity) appear to have considerable generality. Whether they can be accounted for in structural terms is a question still to be answered.

Modes of Resolution

One of the major problems in cognitive consistency theory is that of specifying which mode of resolution an individual will adopt to restore consistency. In the present context the problem is: Which relation will change when the perceived structure is unbalanced? In the case of the communication model, for example, will the receiver change his attitude toward the issue, alter his attitude toward the source, misperceive the stand taken by the communication, etc.?

For each of the three models, one might expect changes to occur in the weaker relations of a perceived structure first, but that such changes will be conditional upon the precise nature of the structure. If a change in a weak relation were to throw a lot of semicycles out of balance, for example, such a change might be resisted (see Figure 8). Some relations in a perceived structure may be more tied to the immediate situation than others (e.g., one's attitudinal relation to a source who is obviously weak and ill-informed) and may be more easily changed than others which are less situation specific and more self-relevant. Furthermore, multiple changes may occur in a perceived structure as well as single changes. For example, a number of changes are allowed for in the congruity model (Osgood & Tannenbaum, 1955). Perhaps also there is a tendency to keep the number of changes needed to restore balance to a minimum and to maximize hedonic gain and minimize hedonic loss (Rosenberg & Abelson, 1960). Abelson (1968a) has succinctly summarized a number of alternative hypotheses concerning modes of resolution.

These comments concern *cognitive adjustments* and that is in line with the main emphasis of the present paper. But it should be evident that overt behavioral adjustments may also occur in addition to or in place of covert adjustments. Confronted by inconsistency, the individual may as a first step seek further information by exploration, communication, or other means and this may lead to modification in existing structures (Berlyne, 1960, 1966; Newcomb, 1959), or he may engage in escape behavior and remove himself from the source of inconsistency. What adjustment is made probably relates to the hierarchy of habits that the individual has established in the course of his developmental history. For some individuals, overt action may be a dominant response without intervening cognitive adjustment; for others, cognitive adjustment may be dominant and may precede overt action. A complete understanding of modes of resolution, therefore, will require specification of the habit hierarchy for individuals in relation to particular situations. The specific adjustments made may also reflect the influence of motivational factors in addition to consistency seeking and inconsistency reduction. A recent analysis of information-seeking behavior by the author has allowed for such extrinsic factors within an expectancy-value paradigm (Feather, 1967a). The whole general question of the relationship of behavioral adjustment to cognitive adjustment is a particularly thorny one in psychology. It may turn out that the relationship is correlative rather than causal (Atkinson, 1969, pp. 123–126).

---

[2] The present models do not include representation of ambivalent relations. A recent attempt has been made by Cartwright and Harary (1970) to treat the problem of ambivalence and indifference.

Individual Differences

The present approach allows for individual differences both in perceived structures and in abstract structures. Subjects may differ in their immediate perceptions of the same situation and in the abstract structures they employ. As already indicated, for some persons the abstract structures may be complex and differentiated, for others simple and undifferentiated. Investigation of individual differences in structures implies the need for a taxonomy of ways in which structures can differ. Zajonc (1968, pp. 320–338) has provided an excellent review of approaches to cognitive organization in terms of undefined elements, attributes, dimensions, categories, concepts, and schemata. This paper will not attempt to cover the same ground. Of special relevance, however, are attempts to formulate structural aspects of cognitive organization (e.g., Scott, 1963, 1969; Zajonc, 1960) employing concepts such as differentiation, complexity, unity, rigidity, centrality, etc.—concepts that are in turn related to Lewin's pioneering attempts to use topology in the representation of a person's life space (Lewin, 1935, 1951). Various ways of distinguishing between structures have also been suggested in the analysis of sociometric structures and communication networks (Bavelas, 1950; Shaw, 1964). One should be able to give a more precise definition of these structural concepts using graph theory, and although some approaches have been made in this direction (e.g., in group structure, Flament, 1963), much more needs to be done. Such translations would have immediate value since a person's conceptual organization has been taken by some as a means of understanding his personality, his cognitive style, his level of psychological development, etc. (e.g., Kelly, 1955; open-mindedness versus closed-mindedness, Rokeach, 1960; concreteness–abstractness, Harvey, Hunt, & Schroder, 1961).

A number of studies have focused on general personality variables such as dogmatism, category width, intolerance of ambiguity, concreteness–abstractness, self-esteem, etc., in relation to tolerance for inconsistency (Glass, 1968), but these studies are beset by difficulties of interpretation (Abelson, 1968b). Thus, in the present context, tests of the recall model have shown no effects of dogmatism or intolerance of ambiguity on the types of argument available in recall (Feather, 1970), yet in another study two personality variables (dogmatism and intolerance of ambiguity) were found to influence preference for information that varied in consistency and novelty (Feather, 1969e). It is likely that such research would be more productive if the personality variables could be closely related to well-defined dimensions of the underlying abstract structures. One obvious dimension already discussed is the sheer size of the structure in terms of the number of semicycles it contains. It is a plausible assumption that individuals who are to maintain relatively complex and differentiated abstract structures containing large numbers of semicycles are more likely to be able to tolerate inconsistencies since, as argued previously, large complex structures are more difficult to balance completely than simple structures that are relatively undifferentiated.

Summary

The present section has been concerned with ways in which the three models might be expanded. The treatment has necessarily been rather summary, but new directions have been suggested that are feasible and that could be pursued in the future to enable an even more powerful representation of cognitive structures concerned with communication, recall, and attribution, and an even better understanding of the various cognitive adjustments and behavioral effects that may occur.

The remaining sections of this article will move away from discussion of the three models to focus on some wider issues. These are concerned with the possibility of generalizing the principle of structural discrepancy to structures relating to the physical world, the problem of development and change in abstract structures, and finally whether there are ways (apart from balance) in which abstract structures might be organized. In conclusion, brief mention will be made of some approaches in the literature that are similar in certain respects to the present one.

ABSTRACT STRUCTURES AND PHYSICAL REALITY

Although previous discussion involving the three models has referred only to situations in which the relevant structures contain personal as well as impersonal elements, it should be apparent that abstract structures may refer both to the physical environment and to the social environment. One's way of structuring the world of physical objects is probably more stable than one's way of structuring social objects, but it is believed that the attempt to resolve structural discrepancy applies to both. In both cases a discrepancy between a perceived structure and an abstract structure may be assumed to lead to pressure to resolve the discrepancy by cognitive and/or behavioral adjustment. The familiar distorted room illusion of Ames is a good example of a case in which a perceived structure is discrepant with an abstract structure that refers to the physical environment (one's conception of the way people usually appear in rooms); so is the perception of a picture hanging askew on the wall (it ought to hang straight). The strange figures that involve contour confusions likewise involve discrepancies between perceived structures and abstract structures based on one's usual experience of contours in the physical world (see the disconcerting example in Figure 10). Hebb's earlier research on the discrepancy between the expected and the strange also falls into this class (Hebb, 1949), since the perceived structure (concerning, for example, the laboratory attendant dressed in unfamiliar clothes) is discrepant with an abstract structure (concerning the way the laboratory attendant usually dresses).

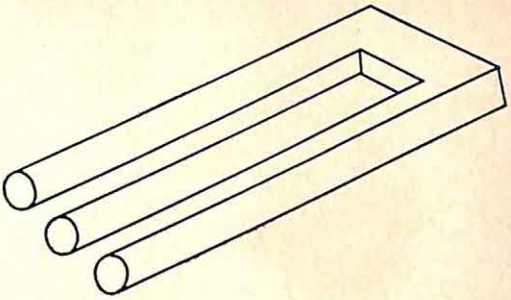


FIG. 10. Perceived structure inconsistent with underlying abstract structure concerning contours.

Such discrepancies create tension and there may be a move to reinterpret information from the physical environment just as one may try to modify perceived structures that are related to social situations and that are discrepant with assumed abstract structures (as in the examples discussed previously relating to communication and attribution). However, as previously indicated, it is probably harder to change one's interpretation of physical reality than to change one's interpretation of social reality (cf. Festinger, 1957, p. 27). Hence, where the discrepant perceived structure refers to the physical environment, direct behavioral adjustments may occur. These adjustments may involve the seeking of new information, particularly when the structural discrepancy is small (as in curiosity behavior, see Berlyne, 1960, 1966), leading to the modification of existing structures or to the development of new ones. Alternatively, active escape from the situation may occur, particularly when the discrepancy is large (as in avoidance of the strange, see Hebb, 1949). Such avoidance would have the effect of preserving existing abstract structures.

DEVELOPMENT AND CHANGE IN ABSTRACT STRUCTURES

Where perceived structures cannot be modified and where the structural discrepancy is consistently experienced, some change in the corresponding abstract structures may develop over time. Such changes occur in the course of cognitive development as the child grows older and as more and more novel and discrepant information is fed into the system (cf. Piaget). The old structures no longer prove adequate and are modified and elaborated to meet the changing circumstances. Hunt (1961, pp. 267–288) has provided an interesting account of accommodative modification and growth in terms of the match between environmental circumstances and existing schemata, using the Piagetian stages of development. He suggests that

> any discrepancy between central processes and circumstance beyond the limits of an organism's capacity for accommodation evokes distress and avoidance, while any discrepancy within the limits of an organism's capacity for accommodation is a source of pleasurable interest or curiosity . . . when circumstance and central process match perfectly, the result is stultifying boredom in which development fails [p. 269].

Such assumptions are obviously very similar to those implied by the work of Hebb (1949) and Berlyne (1960, 1966) and to those presented in the previous section.

The abstract structures involved in the three models that have been described (communication, recall, attribution) are probably well-established ones, end products of a lot of experience in social situations. At a more radical level, change in abstract structures may be forced by an extreme change in perceived structures. In regard to physical reality, the experiments concerning the effects of reversal or distortion of the visual field probably involved such changes (e.g., Kohler, 1962). In regard to social reality, a sudden shift to a hypothetical culture in which success was negatively valued and failure was not a disgrace would probably lead to basic changes in the abstract structures assumed for the attribution model.

Finally, one may deliberately attempt to create new abstract structures by reexposing oneself to past information (if it is possible) and by refusing to interpret it in terms of existing structures that have proved inadequate. Such a process may underly the creation of new scientific theories and perhaps a willingness to dissociate oneself from abstract structures already available.

ORGANIZATION OF ABSTRACT STRUCTURES

In presenting the three models of communication, recall, and attribution earlier, it was assumed that the underlying abstract structures would be organized according to the principle of balance. In these models, both personal as well as impersonal elements were involved, and relations could be attitudinal or unit. It was recognized that in more complex structures, involving large numbers of elements and relations, complete balance may not be possible but that there should be an overall tendency for the structure to be well organized. The present section is concerned with whether or not there are other principles of organization apart from balance that may be applied to abstract structures. Is the balance principle a general one or is it only one of a number of principles underlying cognitive organization?

This question highlights an important issue in cognitive psychology today. While there has been general agreement that cognitive organization occurs (witness the use of terms such as structures, schemata, clusters, chunks, etc.), not enough attention has been paid to how the structures emerge and what principles govern their organization. In the case of structures related to the physical world, the tendency toward formal organization of objects and events on the basis of principles of similarity, temporal and spatial contiguity, closure, symmetry, good continuation, etc., is undoubtedly important and is a major contribution of the Gestalt psychologists, though one might not want to accept their notion of innate antecedents at this stage, but to leave the question open. In his discussion of cognitive structures, Neisser (1967, pp. 290–292) underlines the importance of temporal and spatial dimensions in the way structures are organized. Such dimensions presumably enable the reconstruction of physical and social events in their temporal order as well as in terms of their spatial characteristics. Also, important in regard to abstract structures that are keyed to the physical world are the various *constancies* related to size, shape, color, etc. These groupings and constancies give stability and consistency to experience. From the *empiricist* point of view, they may be seen as summaries or abstractions from continual experience with the physical environment, becoming essential aspects of the programs or abstract structures that underly one's stable view of the physical world (cf. Hebb, 1949). A *nativist* approach (as in Gestalt psychology) would see them as outcomes of innately determined brain processes, determining the form in which information input is cognized (see Neisser, 1967, p. 246).

But what of social reality? Do the organizational principles of Gestalt psychology also apply to the social world? Zajonc (1968, p. 323) has pointed to the dangers in making this assumption. Yet, as used by Heider (1958), the principle of balance is rather similar to the Gestalt principle of *pragnänz*—the emphasis on "good form" and a fieldlike state of equilibrium. Fundamental to this notion is that of a cognitive unit having parts that are of the same *dynamic character* (Heider, 1944, 1946; Zajonc, 1968, p. 339). When this similarity occurs, the parts are grouped together and a steady state exists. But when parts are of different dynamic character, disequilibrium arises and the parts tend to become segregated from each other (Heider, 1946, p. 107). In his cogent discussion of the principle of balance, Zajonc (1968) argues that although forces toward balance have the same character as Gestalt forces toward "good figures" in the perception of forms, the structures considered by Heider have in addition another component, namely, the forces deriving from particular attitudes (see also Heider, 1960, p. 168). According to Zajonc it has

> never been theoretically clear how these attitudinal forces *interact* with the "good figure" forces—whether the former are more or less important, or even whether they have the same dynamic properties as Gestalt forces [p. 341].

The intrusion of these "attitudinal forces" has become evident in a number of studies testing the balance principle. Recently, consideration of the attitudinal relation between two persons (P and O) in a P-O-X structure also involving impersonal object (X), has led Newcomb (1968) to distinguish between balanced, unbalanced, and nonbalanced sets of relations. A balanced set of relations is one which a person accepts as it is; an unbalanced set of relations is one that he attempts to modify; a nonbalanced set of relations is one toward which he has a clear lack of preference for accepting or modifying the set of relations. According to Newcomb, whenever the PO relation is negative (one person dislikes the other) the triad is *nonbalanced* and is often accompanied by indifference and low engagement. Otherwise his definitions are the same as Heider's. Newcomb's discussion also indicates that the original definition of balance may be oversimplified and his review of the research evidence tends to support his view (see also Zajonc, 1968, Table 1). Thus, the balance principle is coming under increasing attack not only on theoretical grounds, but also in relation to how far it can handle all of the research evidence. It may end up as only one of a number of factors influencing cognitive adjustments.

The use of the balance principle in the three models described previously should not be taken as endorsement of Gestalt principles of *pragnänz,* closure, etc. Such principles have not fared well as unitary explanations (for a recent review of studies concerning closure see Holmes, 1968), and the underlying assumption of innate determinants, particularly of spatial organization, is a hotly disputed one. Instead, in the three models described in the present article, the balance principle is used in a *definitional* sense to indicate what is meant by consistent and inconsistent sets of relations (something that is often left vague in alternative consistency theories—e.g., dissonance theory). In a very broad sense, each of the three models may be seen as presenting *a structural set of rules which are organized in a consistent manner (balanced) and which specify the psycho-logic of how sets of elements are associated and evaluated*. One way of interpreting the balanced abstract structures, therefore, is in terms of systems of rules that constitute internally consistent theories about reality. For example, on the basis of one's own experience in task situations, one might formulate the general rule that unexpected success or unexpected failure at a task will usually be caused by variable characteristics of the person or the situation (e.g., by an unusual increase or decrease in momentary effort or by good or bad luck), whereas expected outcomes will be caused by stable dispositional properties of the person or situation (e.g., by the presence or absence of ability or the ease or difficulty of the task). This general rule is captured in the abstract structures of the attribution model insofar as the person is concerned. Similarly, on the basis of past experience in communication situations, one might formulate the general rule that people whom one likes usually have attitudes toward issues similar to one's own attitudes, and that their communications tend to be consistent with these attitudes. This general rule is captured in those abstract structures of the communication model in which the attitudinal relations between source and receiver are positive. Thus, the balance principle enables one to structure or organize sets of general rules in a consistent manner and to predict the direction a particular rule will take within the total system of rules.

In an important sense, therefore, the underlying abstract structures involved in the representation of social reality have a *normative* function. They specify sets of rules making up internally consistent theories about what *ought* to be the case in any given social situation that has been experienced over time. It is likely that the normative aspect of abstract structures also extends to structures concerning physical reality. Constancies of size and shape, for example, may be considered to involve notions of what ought to be the case in the physical environment. This quite general interpretation of abstract structures recognizes that one of the compelling characteristics of individuals is that they follow rules and are guided by normative principles (Peters, 1969; Sherif, 1936).

Why are the rules balanced or consistent in the three models presented? The answer to this question is beyond the scope of the present article. Yet, as indicated above, we do not want to prejudge the issue by appealing to innately determined principles of organization, though innate determinants may be part of the answer. Zajonc (1968) indicates that

> Whether or not these principles can be extended to refer also to semantic or social proximity, semantic or social similarity, semantic or social continuation, and semantic or social closure is an empirical question which has not yet been answered [p. 342].

As emphasized previously, modes of organization may be learned in the course of an individual's adaptation to a changing world to give meaning, continuity, and stability to variable information input, and such learning may be an essential part of early experience (Hebb, 1949). In addition, there are demands on a person from an early age toward behaving consistently, and in time, consistency may acquire incentive value in its own right, although one would expect individual differences to occur in such learning (Feather, 1967b). Finally, the real world may be constructed in a consistent way in the sense that, for example, friends do tend to like the same things, people do tend to like the things they own, etc. The balanced organization of abstract structures keyed to social contexts may therefore reflect social reality. Certainly, if one relates abstract structures to sets of rules organized in a consistent way according to the balance principle, then these normative principles would be influenced by what *usually* happens in real social situations, since abstractions should incorporate modal events.

Some of the ideas expressed above have their counterparts in the literature of consistency theory, though these related approaches do not make the distinction between perceived structures and abstract structures and have little to say about the long-term effects of information processing in the formation of abstract structures. For example, some theories have specifically looked at organization in terms of rules of implication, emphasizing a sort of psychological rationality in man and indicating principles of "psycho-logic" by which consistency is achieved (Abelson, 1968c; Abelson & Rosenberg, 1958; McGuire, 1968; Rosenberg & Abelson, 1960). Jordan (1968) has related imbalance to conflicting implications about unit formation among sets of elements and relations and argues that "*p*U*o* and *p*-U*o* is the canonic form for any form of imbalance, no matter how many entities are involved [p. 175]." In other cases more emphasis has been placed on conceptual good figures in line with the Gestalt tradition that the configurational laws of perception apply to cognition generally (Koffka, 1935). Thus, De Soto (1960, 1961) has shown that there is a pervasive preference for linear orderings when people assign orders and that this preference may be associated with an underlying conceptual good figure or schema (see also De Soto & Albrecht, 1968). More recently, Davis (1967) has argued for a general principle of clusterability as a basis for organization, particularly within sociometric structures. The balance principle implies the formation of two clusters or cliques linking elements, with positive relations within each clique and negative relations between cliques (see the structure theorem, Cartwright & Harary, 1956). Davis has suggested that more than two cliques may form under certain conditions that can be specified mathematically and that balance may be a special case of clusterability. Research by De Soto, Henley, and London (1968) does suggest that clusterable structures (more than two cliques) are as easily learned as balanced structures, but such research is still in its infancy (see also Peay, 1970). These different contributions indicate that the balance principle may be only one likely principle underlying organization in abstract structures and that at the present stage, one should not exclude the possibility that a number of principles may be found to exist, some more general than others. The balance principle, however, does seem to be a viable one in relation to the three models discussed in the present article, enabling a precise definition of how relations within the abstract structures are organized in regard to their signs. But, as previously indicated, it is coming under increasing attack as a unitary principle and in relation to the empirical evidence. It probably needs to be supplemented by consideration of other factors (cf. the previous discussion of expanded models and of specific attitudinal effects; Newcomb, 1968; Zajonc, 1968). As a principle, it is more appropriate to abstract structures keyed to social reality where both attitudinal and unit relations are involved.[3] Moreover, in time it may be superseded by more general principles such as clusterability.

Finally, it should not be forgotten that the perceived structure is assumed to be organized too and that the processes involved in this organization also need to be understood. In most cases, particularly with respect to the physical environment, perceived structures and abstract structures tend to correspond. The interesting cases emerge when organization in the perceived structure is imperfect or abnormal (as in Figure 10 or in an unbalanced communication situation), since there is then a discrepancy between the perceived structure and an underlying abstract structure. Consideration of the effects of these structural discrepancies has been a major thrust of the present article.

DISCREPANCY THEORY: RELATED APPROACHES

The general paradigm involved in the present approach is one that focuses on structural discrepancies and their cognitive and behavioral effects. The use of a concept of discrepancy occurs in other theories in psychology and some of these have already been mentioned (e.g., Hebb's analysis of the fear of the strange, Berlyne's discussion of curiosity, Hunt's account of the match between environmental circumstances and existing schemata). The general approach of the consistency theories is also somewhat similar. Newcomb (1959), for example, has included a concept of discrepancy in his analysis of the strain toward symmetry, and Aronson (1968) has considered dissonance theory in terms of confirmation or disconfirmation of self-relevant expectancies. The earlier contributions of Festinger (1950, 1954) concerning informal social communication and social comparison also involved concepts of discrepancy. These approaches, however, have not distinguished between abstract structures and perceived structures. The following related approaches may also be briefly noted: the emphasis on general schemata that are built up over time and tested against reality (Bartlett, 1932; Piaget, 1966), spontaneous changes in recall in terms of underlying ideal structures (Koffka, 1935), generalized expectancies or "cognitive maps" that can be confirmed or disconfirmed by experience (Tolman, 1958), neural organizations that become progressively refined as learning proceeds (Hebb, 1949), adaptation levels that are based on previous experience (Helson, 1964), discrepancies between perception and adaptation levels that underly the learning of motives (McClelland, Atkinson, Clark, & Lowell, 1953, Ch. 2), disparities in psychological events that define motives (Peak, 1968), images and plans (analogous to computer programs) that guide the form that behavior takes (Miller, Galanter, & Pribram, 1960), categories or concepts that are refined on the basis of experience and to which information input is matched in some sense so as to achieve stability in perception and thought (Bruner, 1957; Brunswik, 1939; Harvey et al., 1961), and basic linguistic structures that enable us to understand the generation of sentences (Chomsky, 1957, 1965).

---

[3] The present discussion has not included an analysis of the types of structures considered by Piaget. The three models have been concerned with structures containing both persons and things as elements and with relations that include evaluations as well as unit-forming tendencies. The mathematics is in terms of graph theory. Piaget's structures focus more on a psychologic involving lattice theory and group theory. But the Piagetian emphasis on adaptation and organization is basic to the present formulation.

All of these approaches are similar in some respects to the present one though they differ in the degree to which the structural concepts involved in them are articulated in formal terms. For example, even though the concept of expectancy is a structural concept involving the basic idea that one comes to learn the implications of a response (e.g., If in this situation (S) I perform this response (R), then outcome (X) will happen; see MacCorquodale & Meehl, 1953), very little has been done to represent expectancies and the discrepancy between expectancy and present experience using structural models incorporating mathematical treatments of structures. And the same criticism applies to most of the other concepts involved in the above list.

CONCLUDING REMARKS

The concepts employed in the present article obviously have wide generality. They focus on the individual's attempt to achieve some degree of constancy or stability as he processes information from the physical and social world, and the consequences that follow when this constancy is violated. In a very general sense, the abstract structures involve internally consistent systems of rules or theories about what *ought* to be the case in the physical and social environment. The development of organized abstract structures appears to be a necessary part of an individual's adaptation to a complex world.[4]

The generality of these concepts and the fact that similar approaches have appeared in various areas of psychology suggest the emergence of a basic paradigm (Kuhn, 1966). It is a paradigm that forces one to come to grips with the concept of structure, to consider different ways in which structures may be organized and ways in which structures may be discrepant. The present paper has shown that it is possible to deal with these questions mathematically (albeit in a simple way), thereby providing a clear definition of structural organization, and that it is also possible to determine empirically what the perceived and abstract structures are and the effects that occur when there is discrepancy between them. Much remains to be done, however, in regard to the mapping of cognitive structures, the identification of modes of organization, and the determination of degrees of structural discrepancy. Further development in this area is an important task for psychology since structural concepts lie at the heart of so much of psychological theory.

---

[4] The detailed elaboration of this idea in relation to development is, of course, a major contribution of Piagetian psychology (Piaget, 1966).

REFERENCES

ABELSON, R. P. A summary of hypotheses on modes of resolution. In R. P. Abelson, E. Aronson, W. J. McGuire, T. M. Newcomb, M. J. Rosenberg, & P. H. Tannenbaum (Eds.), *Theories of cognitive consistency: A sourcebook.* Chicago: Rand McNally, 1968. (a)

ABELSON, R. P. Comment: Uncooperative personality variables. In R. P. Abelson, E. Aronson, W. J. McGuire, T. M. Newcomb, M. J. Rosenberg, & P. H. Tannenbaum (Eds.), *Theories of cognitive consistency: A sourcebook.* Chicago: Rand McNally, 1968. (b)

ABELSON, R. P. Psychological implication. In R. P. Abelson, E. Aronson, W. J. McGuire, T. M. Newcomb, M. J. Rosenberg, & P. H. Tannenbaum (Eds.), *Theories of cognitive consistency: A sourcebook.* Chicago: Rand McNally, 1968. (c)

ABELSON, R. P., & ROSENBERG, M. J. Symbolic psycho-logic: A model of attitudinal cognition. *Behavioral Science,* 1958, **3**, 1–13.

ARONSON, E. Dissonance theory: Progress and problems. In R. P. Abelson, E. Aronson, W. J. McGuire, T. M. Newcomb, M. J. Rosenberg, & P. H. Tannenbaum (Eds.), *Theories of cognitive consistency: A sourcebook.* Chicago: Rand McNally, 1968.

ATKINSON, J. W. Change of activity: A new focus for the theory of motivation. In T. Mischel (Ed.), *Human action.* New York: Academic Press, 1969.

ATKINSON, J. W., & FEATHER, N. T. (Eds.). *A theory of achievement motivation.* New York: Wiley, 1966.

BARTLETT, F. C. *Remembering: A study in experimental and social psychology.* Cambridge: Cambridge University Press, 1932.

BAVELAS, A. Communication patterns in task-oriented groups. *Journal of the Acoustical Society of America,* 1950, **22**, 725–730.

BERLYNE, D. E. *Conflict, arousal, and curiosity.* New York: McGraw-Hill, 1960.

BERLYNE, D. E. Curiosity and exploration. *Science,* 1966, **153**, 25–33.

BOUSFIELD, W. A. The occurrence of clustering in the recall of randomly arranged associates. *Journal of General Psychology,* 1953, **49**, 229–240.

BRUNER, J. S. On perceptual readiness. *Psychological Review,* 1957, **64**, 123–152.

BRUNSWIK, E. The conceptual focus of some psychological systems. *Journal of Unified Science,* 1939, **8**, 36–49.

CARTWRIGHT, D., & HARARY, F. Structural balance: A generalization of Heider's theory. *Psychological Review,* 1956, **63**, 277–293.

CARTWRIGHT, D., & HARARY, F. Ambivalence and indifference in generalizations of structural balance. *Behavioral Science,* 1970, **15**, 497–513.

CHOMSKY, N. *Syntactic structures.* 's Gravenhage, Netherlands: Mouton, 1957.

CHOMSKY, N. *Aspects of the theory of syntax.* Cambridge, Mass.: M.I.T. Press, 1965.

DAVIS, J. A. Clustering and structural balance in graphs. *Human Relations,* 1967, **20**, 181–187.

DEESE, J. *The structure of associations in language and thought.* Baltimore: Johns Hopkins Press, 1965.

DE SOTO, C. B. Learning a social structure. *Journal of Abnormal and Social Psychology,* 1960, **60**, 417–421.

DE SOTO, C. B. The predilection for single orderings. *Journal of Abnormal and Social Psychology,* 1961, **62**, 16–23.

DE SOTO, C. B., & ALBRECHT, F. Conceptual good figures. In R. P. Abelson, E. Aronson, W. J. McGuire, T. M. Newcomb, M. J. Rosenberg, & P. H. Tannenbaum (Eds.), *Theories of cognitive consistency: A sourcebook.* Chicago: Rand McNally, 1968.

DE SOTO, C. B., HENLEY, N. M., & LONDON, M. Balance and the grouping schema. *Journal of Personality and Social Psychology,* 1968, **8**, 1–7.

FEATHER, N. T. Subjective probability and decision under uncertainty. *Psychological Review,* 1959, **66**, 150–164.

FEATHER, N. T. A structural balance model of communication effects. *Psychological Review,* 1964, **71**, 291–313.

FEATHER, N. T. A structural balance analysis of evaluative behavior. *Human Relations,* 1965, **18**, 171–186. (a)

FEATHER, N. T. Reactions to communications under conditions of source responsibility and source coercion. *Australian Journal of Psychology,* 1965, **17**, 179–194. (b)

FEATHER, N. T. The prediction of interpersonal attraction: Effects of sign and strength of relations in different structures. *Human Relations,* 1966, **19**, 213–237.

FEATHER, N. T. An expectancy-value model of information-seeking behavior. *Psychological Review,* 1967, **74**, 342–360. (a)

FEATHER, N. T. A structural balance approach to the analysis of communication effects. In L. Berkowitz (Ed.), *Advances in experimental social psychology.* Vol. 3. New York: Academic Press, 1967. (b)

FEATHER, N. T. Effects of institutional affiliation and attitude discrepancy on evaluation of communications and interpersonal attraction. *Human Relations,* 1967, **20**, 101–120. (c)

FEATHER, N. T. Attitude and selective recall. *Journal of Personality and Social Psychology,* 1969, **12**, 310–319. (a)

FEATHER, N. T. Attribution of responsibility and valence of success and failure in relation to initial confidence and task performance. *Journal of Personality and Social Psychology,* 1969, **13**, 129–144. (b)

FEATHER, N. T. Cognitive differentiation, attitude strength, and dogmatism. *Journal of Personality,* 1969, **37**, 111–126. (c)

FEATHER, N. T. Differentiation of arguments in relation to attitude, dogmatism, and intolerance of ambiguity. *Australian Journal of Psychology,* 1969, **21**, 21–29. (d)

FEATHER, N. T. Preference for information in relation to consistency, novelty, intolerance of ambiguity, and dogmatism. *Australian Journal of Psychology,* 1969, **21**, 235–249. (e)

FEATHER, N. T. Balancing and positivity effects in social recall. *Journal of Personality,* 1970, **38**, 602–628.

FEATHER, N. T., & ARMSTRONG, D. J. Effects of variations in source attitude, receiver attitude and communication stand on reactions to source and content of communications. *Journal of Personality,* 1967, **35**, 435–455.

FEATHER, N. T., & JEFFRIES, D. G. Balancing and extremity effects in reactions of receiver to source and content of communication. *Journal of Personality,* 1967, **35**, 194–213.

FEATHER, N. T., & SIMON, J. G. Attribution of responsibility and valence of outcome in relation to initial confidence and success and failure of self and other. *Journal of Personality and Social Psychology,* 1971, **18**, 173–188. (a)

FEATHER, N. T., & SIMON, J. G. Causal attributions for success and failure in relation to expectations of success based upon selective and manipulative control. *Journal of Personality,* 1971, in press. (b)

FESTINGER, L. Informal social communication. *Psychological Review,* 1950, **57**, 271–282.

FESTINGER, L. A theory of social comparison processes. *Human Relations,* 1954, **7**, 117–140.

FESTINGER, L. *A theory of cognitive dissonance.* Evanston, Ill.: Row, Peterson, 1957.

FLAMENT, C. *Applications of graph theory to group structure.* Englewood Cliffs, N. J.: Prentice-Hall, 1963.

GLASS, D. C. Theories of consistency and the study of personality. In E. F. Borgatta & W. W. Lambert (Eds.), *Handbook of personality theory and research.* Chicago: Rand McNally, 1968.

HARARY, F., NORMAN, R. Z., & CARTWRIGHT, D. *Structural models: An introduction to the theory of directed graphs.* New York: Wiley, 1965.

HARVEY, O. J., HUNT, D. E., & SCHRODER, H. M. *Conceptual systems and personality organization.* New York: Wiley, 1961.

HEBB, D. O. *The organization of behavior.* New York: Wiley, 1949.

HEIDER, F. Social perception and phenomenal causality. *Psychological Review,* 1944, **51**, 358–374.

HEIDER, F. Attitudes and cognitive organization. *Journal of Psychology,* 1946, **21**, 107–112.

HEIDER, F. *The psychology of interpersonal relations.* New York: Wiley, 1958.

HEIDER, F. The gestalt theory of motivation. *Nebraska Symposium on Motivation,* 1960, **8**, 145–172.

HELSON, H. *Adaptation-level theory; An experimental and systematic approach to behavior.* New York: Harper, 1964.

HOLMES, D. S. Search for "closure" in a visually perceived pattern. *Psychological Bulletin,* 1968, **70**, 296–312.

HUNT, J. McV. *Intelligence and experience.* New York: Ronald Press, 1961.

JORDAN, N. Cognitive balance as an aspect of Heider's cognitive psychology. In R. P. Abelson, E. Aronson, W. J. McGuire, T. M. Newcomb, M. J. Rosenberg, & P. H. Tannenbaum (Eds.), *Theories of cognitive consistency: A sourcebook.* Chicago: Rand McNally, 1968.

KELLY, G. A. *The psychology of personal constructs.* New York: Norton, 1955.

KOFFKA, K. *Principles of gestalt psychology.* New York: Harcourt, Brace, 1935.

KOHLER, I. Experiments with goggles. *Scientific American,* 1962, **206**, 62–85.

KUHN, T. S. *The structure of scientific revolutions.* Chicago: University of Chicago Press, 1966.

LEWIN, K. *Dynamic theory of personality.* New York: McGraw-Hill, 1935.

LEWIN, K. *Field theory in social science.* New York: Harper, 1951.

MCCLELLAND, D. C., ATKINSON, J. W., CLARK, R. A., & LOWELL, E. L. *The achievement motive.* New York: Appleton-Century-Crofts, 1953.

MACCORQUODALE, K., & MEEHL, P. E. Preliminary suggestions as to formalization of expectancy theory. *Psychological Review,* 1953, **60**, 55–63.

MCGUIRE, W. J. Theory of the structure of human thought. In R. P. Abelson, E. Aronson, W. J. McGuire, T. M. Newcomb, M. J. Rosenberg, & P. H. Tannenbaum (Eds.), *Theories of cognitive consistency: A sourcebook.* Chicago: Rand McNally, 1968.

MILLER, G. A., GALANTER, E., & PRIBRAM, K. H. *Plans and the structure of behavior.* New York: Holt, 1960.

NEISSER, U. *Cognitive psychology.* New York: Appleton-Century-Crofts, 1967.

NEWCOMB, T. M. Individual systems of orientation. In S. Koch (Ed.), *Psychology: A study of a science.* Vol. 3. New York: McGraw-Hill, 1959.

NEWCOMB, T. M. Interpersonal balance. In R. P. Abelson, E. Aronson, W. J. McGuire, T. M.

Newcomb, M. J. Rosenberg, & P. H. Tannenbaum (Eds.), *Theories of cognitive consistency: A sourcebook.* Chicago: Rand McNally, 1968.

OESER, O. A., & HARARY, F. A mathematical model for structural role theory: I. *Human Relations,* 1962, **15**, 89–109.

OESER, O. A., & HARARY, F. A mathematical model for structural role theory: II. *Human Relations,* 1964, **17**, 3–17.

OSGOOD, C. E., & TANNENBAUM, P. H. The principle of congruity in the prediction of attitude change. *Psychological Review,* 1955, **62**, 42–55.

PEAK, H. Activation theory. In R. P. Abelson, E. Aronson, W. J. McGuire, T. M. Newcomb, M. J. Rosenberg, & P. H. Tannenbaum (Eds.), *Theories of cognitive consistency: A sourcebook.* Chicago: Rand McNally, 1968.

PEAY, E. R., JR. Extensions of clusterability to quantitative data with an application to the cognition of political attitudes. Unpublished doctoral dissertation, University of Michigan, 1970.

PETERS, R. S. Motivation, emotions, and the conceptual schemes of common sense. In T. Mischel (Ed.), *Human action: Conceptual and empirical issues.* New York: Academic Press, 1969.

PIAGET, J. *The origin of intelligence in the child.* London: Routledge & Kegan Paul, 1966.

ROKEACH, M. *The open and closed mind.* New York: Basic Books, 1960.

ROSENBERG, M. J. Hedonism, inauthenticity and other goads toward expansion of consistency theory. In R. P. Abelson, E. Aronson, W. J. McGuire, T. M. Newcomb, M. J. Rosenberg, & P. H. Tannenbaum (Eds.), *Theories of cognitive consistency: A sourcebook.* Chicago: Rand McNally, 1968.

ROSENBERG, M. J., & ABELSON, R. P. An analysis of cognitive balancing. In C. I. Hovland & M. J. Rosenberg (Eds.), *Attitude organization and change.* New Haven, Conn.: Yale University Press, 1960.

SCOTT, W. A. Cognitive complexity and cognitive balance. *Sociometry,* 1963, **26**, 66–74.

SCOTT, W. A. Structure of natural cognitions. *Journal of Personality and Social Psychology,* 1969, **12**, 261–278.

SHAW, M. E. Communication networks. In L. Berkowitz (Ed.), *Advances in experimental social psychology.* Vol. 1. New York: Academic Press, 1964.

SHERIF, M. *The psychology of social norms.* New York: Harper, 1936.

SHUELL, T. J. Clustering and organization in free recall. *Psychological Bulletin,* 1971, in press.

SIMON, J. G. Attribution of responsibility for own and other's task outcome in relation to expectation of success and task outcome for self and other. Unpublished honours thesis, Flinders University of South Australia, 1969.

TOLMAN, E. C. *Behavior and psychological man.* Berkeley: University of California Press, 1958.

ZAJONC, R. B. The process of cognitive tuning in communication. *Journal of Abnormal and Social Psychology,* 1960, **61**, 159–167.

ZAJONC, R. B. Cognitive theories in social psychology. In G. Lindzey & E. Aronson (Eds.), *The handbook of social psychology.* Vol. 1. Reading, Mass.: Addison-Wesley, 1968.

(Received October 22, 1970)

Psychological Review
1971, Vol. 78, No. 5, 380–390

THE DUPLICATION THEOREM OF SOCIAL RELATIONSHIPS AS TESTED IN THE GENERAL POPULATION

WALTER TOMAN[1]

University of Erlangen-Nürnberg, Germany

The effects of certain characteristics of a person's family constellation on his social behavior, on his choices of partners for life, and on the family he founds have been tested with two samples of the general population. The "duplication theorem," stating that new social relationships tend to have more chances of permanence and success the closer they resemble earlier and earliest (intrafamilial) social relationships, has been confirmed on a larger scale.

Early social relationships have bearing on later social relationships. This thesis has been advanced by Freud (1953) who claimed explicitly that a person's relations to his parents while a child affect his psychosexual development and, ultimately, his adult heterosexual relationships. Adler (1920) questioned what difference the presence of siblings would make for a person's character development and adjustment in later life. He was not very systematic and described only a few sibling positions and their effects. Segments of such effects have been sporadically studied by others (e.g., Koch 1955, 1956, 1960; Krout 1939; Schachter 1964; Schoonover 1959; Warren 1966).

The author's thesis is the following: Sibling positions tend to determine role preferences in social contacts and in enduring peer relationships outside the family. Friends and partners in love and/or marriage will tend to get along with one another better and for longer times, other things being equal, if their role preferences are compatible. But when are they compatible?

Of many possible aspects by which such compatibility of role preferences may be defined, what was thought to be the simplest and most objective aspects was chosen: the age order and the sex distribution among a person's siblings.

If a sibling configuration consists of n children, there are 2^n different patterns in which it may appear if only sex and age order are being varied. A person chosen randomly can hold any of $n2^{n-1}$ sibling positions. Holding a given of n possible positions in his sibling configuration, his n–1 siblings could be any of 2^{n-1} combinations of brothers and sisters. For the sake of simplicity, let us inspect what this means when $n = 2$.

There are four different patterns of sibling configurations: two brothers, two sisters, a brother and a sister, a sister and a brother. A given person coming from a two-children configuration, if male, may be the older or the younger brother either of a brother or of a sister; if female, she may likewise be the older or the younger sister either of a brother or of a sister.

Sibling positions have ordinarily been maintained over many years in the context of the original family. They are a person's earliest, most regular, and inescapable peer experiences. Social contacts with peers in new settings such as playground, church, school, clubs, etc., may be patterned after them.

Each of the above mentioned eight basic sibling positions implies a peer role distinct from those of all other sibling positions. One may assume, on average, that the oldest brother of a brother—b(b)—is used to a senior position among peers and to contacts with a peer of the same sex only; that the oldest brother of a sister—b(s)—is also used to a senior role, but vis-à-vis a person of the opposite sex only; that the younger brother of a brother—(b)b—is accustomed to be led and protected by a male; and the younger brother of a sister—(s)b—by a female.

---

[1] Requests for reprints should be sent to Walter Toman, Psychologisches Institut der Universität Erlangen-Nürnberg, 8520 Erlangen, Bismarckstrasse 1, Germany.

Likewise, the older sister of a sister—s(s)—is used to a senior position among peers of the same sex; the older sister of a brother—s(b)—to a senior position among peers of the opposite sex; the younger sister of a sister—(s)s—may, on the average, be assumed to expect some kind of a leadership and protection from a peer of the same sex; the younger sister of a brother—(b)s—from a peer of the opposite sex.

All of them would tend to be unaccustomed to assuming opposite roles, respectively, or to maintaining contacts with a peer of a sex other than that of their respective sibling. Such an assumption would be plausible by common sense, but is also compatible with learning theory (e.g., Hull, 1943) or psychoanalytic theory (e.g., Freud, 1953). Most generally stated, experiences of a certain type generate preferences for similar experiences rather than different ones. This would at least hold for positive experiences. Negative experiences would create avoidance preferences, but if avoidance is not really possible in the primary milieu, if the brother or the sister are around to stay and cannot be escaped, as seems to be the case in most families, at least practically speaking, one would expect even such negative experiences with siblings to generate certain preferences for similar experiences and similar long-range relationships. As a matter of fact, so-called negative experiences with siblings and other family members are frequently embedded in a context of relative social comfort and positive attachments which are rarely outweighed by the negative experience. Even if they are, a child usually has no other option during his early and most formative years than to put up with them.

It is also consistent with common sense, learning theory, and psychoanalytic theory that generalizations to new social contexts (such as playground, school, professional relationships, friends) occur automatically. Corrections of such generalized expectations through new experience are possible, to be sure, but those expectations tend also to codetermine new experience, the more so, the more pervasive and exclusive of others the original experiences have been.

One of the new contexts to which the original experiences may be generalized are love relationships and marriage. Sibling roles learned in the original family may be conducive to different degrees of compatibility among love and marriage partners. The following matches would be considered compatible by the types of peer experiences that the partners gathered in their original families: b(s)/(b)s, which means the older brother of a sister married to the younger sister of a brother; and (s)b/s(b), which means the younger brother of a sister married to the older sister of a brother. The reasons: In the first match, both partners get what they were used to in their original families. He can lead and protect a female peer. She can accept to be led and protected by him. This is true of the second match as well, although under different auspices: He gets leadership and care from his wife, she has somebody to guide and care for. In both matches the partners duplicate for one another a former sibling. The partners are maximally compatible by their sibling positions. Their relationship is one of role complementarity.

The following two pairs would, on the other hand, be judged as incompatible relationships: b(b)/s(s), which means the older brother of a brother married to the older sister of a sister; and (b)b/(s)s, which means the younger brother of a brother married to the younger sister of a sister. The reasons: In the first match, neither the man nor the woman are used to life with a peer of the opposite sex by their sibling experiences, and both of them have the same age rank. Both are seniors, probably aspiring to lead and protect the other without finding the other inclined to oblige. In the second match, both are supposedly equally unprepared for life with a person of the opposite sex, and both are juniors. Both would tend to expect to be led and/or protected without finding such leadership or protection with their partner. In both matches the partners fail "maximally" to duplicate for one another a former sibling. The partners are mini-

mally compatible by their sibling positions. Their relationship is the opposite of role complementarity. By sibling experience they have a rank and a sex conflict.

This, in essence, illustrates the "duplication theorem" which states more generally and more inclusively: Those of a person's extrafamilial social relationships have better chances of permanence and success that resemble more closely than others the earlier and earliest (intrafamilial) social relationships that the person has experienced in his life (Toman 1959a, 1959b, 1959c, 1962a).

There are, of course, intermediate degrees of complementarity (e.g., an older brother of both a sister and a brother married to the middle sister of two brothers: b(sb)/(b)s(b)), rank conflicts without sex conflicts (e.g., the older brother of a sister married to the older sister of a brother: b(s)/s(b)), and sex-conflicts without rank conflicts (e.g., the older brother of a brother married to the younger sister of a sister: b(b)/(s)s). A rank order of the severity of conflict of sibling positions of love or marriage partners has been established as part of the study to be reported. It is too cumbersome to describe in detail, although a computer program has been written to distinguish nine such degrees of complementarity (even 12, if matches with single children among the partners were included). Suffice it here to make and explain the cruder distinctions upon which the statistical tests to be reported were based.

Matches of *high complementarity* of sibling positions of partners comprise: (*a*) Complete complementarity: Both partners have siblings of the opposite sex only and complement each other by sibling-age-rank; for example, b(ss)/(b)s or (ss)b/s(bbb), which means the older brother of two sisters married to the younger sister of a brother, or the younger brother of two sisters married to the oldest sister of three brothers, respectively.

(*b*) Partial complementarity: One or both partners have siblings of both sexes, but there is at least one sibling in the original family of each partner who duplicates the age-rank and sex of the spouse; for example, b(sb)/(bsb)s, which means the older brother of a sister and a brother married to the youngest sister of two older brothers and one older sister (the husband has a younger sister among his siblings and the wife two older brothers among hers); or (s)b(s)/s(b), which means the middle brother of two sisters married to the older sister of a brother (among the two sibling relationships of the husband, one sibling, his older sister, duplicates the sibling role of his spouse).

Matches of *medium to low complementarity* of sibling positions of partners comprise all other types except those in which single children are among the partners. There are rank conflicts without sex conflicts (e.g., b(ss)/s(b); both partners are the oldest, but both have siblings of the opposite sex and can thus be assumed to be used to life with a peer of the opposite sex; or b(ss)/(s)s(b), which may be considered a milder rank conflict than the first one, since the wife has an older sister and is thus partly accustomed to the role of a junior, if a junior to a female). There are sex conflicts without rank conflicts (e.g., (b)b/s(ss)): she, an older sister of two sisters, would tend to offer guidance and care, and he, the younger brother of a brother, would be willing to accept that, even though they have not learned in their original families to play those parts vis-à-vis a person of the opposite sex. Finally there are rank plus sex conflicts (e.g., b(bb)/s(s)), the older brother of two brothers married to the older sister of a sister, or (bbb)b/(ss)s, the youngest of four brothers married to the youngest of three sisters, or (b)b(b)/s(ss), the middle one of three brothers matched with the oldest of three sisters; the sibling-role conflict of the last couple would be considered slightly milder, that is, something less than a complete rank plus sex conflict, since the husband has had an older sibling too, not only a younger one, and could be expected to find leadership by the partner at least partly acceptable.

Matches with single children among the partners are considered to represent the *lowest degree* of sibling role *complementarity*. The single child has supposedly had no peer experience at all in his original family. Yet if he hits upon a marriage partner who has had a sibling of the opposite sex (e.g., b/s(b)), their match would be deemed somewhat better than a match between a single child and a partner with siblings of the same sex only (e.g., b(bb)/s) or the match of two single children (b/s). The assumption is that the partner with an early peer experience can help the single child better than can another single child to fit into the current peer relationship: marriage.

Repeating, three degrees of complementarity of sibling positions of (marriage) partners have been distinguished: high complementarity (matches with complete or partial complementarity of sibling positions of partners), medium to low complementarity (matches with partial or full rank, sex, or rank plus sex conflict of sibling positions of partners), and lowest complementarity (matches with at least one partner being a single child).

TESTS OF THE "DUPLICATION THEOREM"

One critical test of the duplication theorem would be a comparison of divorced marriages versus intact marriages. If the theorem has validity, divorced marriages should show a higher frequency of rank and/or sex conflict between the sibling positions of the partners than intact marriages.

This was first tested with two groups of divorced and intact marriages (Toman 1962b). The groups were matched by socioeconomic status (education, approximate income, ethnic origin, and age). Pairs with partners who were single children were eliminated. Intact marriages were to have lasted for at least 10 years and to have led to at least two children.

Comparison of the compatibility of their sibling positions showed that of 16 married couples, 12 displayed high complementarity and 4 medium to low complementarity, whereas of the 16 divorced couples, 1 showed high complementarity and 15 medium to low complementarity. As mentioned before there were no single children among the couples. This distribution of frequencies tested with the Fisher-Yates correction formula is highly significant ($\chi^2 = 12.0$, $df = 1$, $p < .001$). In other words, complementarity of sibling positions of partners seems clearly to bear on the outcome of marriage. Low complementarity increases the likelihood of eventual divorce.

One might suspect that these effects could be attributable not so much to complementarity or noncomplementarity of sibling positions of partners, but merely to the number of siblings the partners had. Persons coming from larger families are less likely to divorce each other than are those coming from small families. While this aspect, number of siblings, would be one of several other aspects by which the "duplication theorem" can be tested (and has been tested successfully: see Toman, Preiser, Gasch, & Plattig, 1967), the influence of mere number of siblings regardless of considerations of age-rank and sex of siblings would be somewhat more trivial and should be tested separately. It is logically evident that partners coming from a larger configuration of siblings are more likely than those coming from smaller configurations to have had the "right" sibling among them: one of the opposite sex and of an age-rank partly or fully identical with the spouse's age-rank among his/her siblings.

The samples of 16 divorced and 16 intact marriages showed no significant difference in numbers of siblings in their original families, although a tendency in this direction was apparent (see Table 1).

As can be seen, partners of divorced marriages come from somewhat smaller families (i.e., from families with an "average" of 2.31 and 2.50 children for male and female partners, respectively) than did the married

TABLE 1

NUMBER OF PARTNERS OF DIVORCED AND OF MARRIED COUPLES IN RELATION TO NUMBER OF THEIR SIBLINGS

Status of partner	Male partner[a]		Female partner[b]	
	One sibling	Two or more siblings	One sibling	Two or more siblings
Divorced	12	4	9	7
Married	10	6	6	10

[a] $\chi^2 = .14$, $df = 1$, $p < .75$.
[b] $\chi^2 = .67$, $df = 1$, $p < .50$.

partners (with an "average" of 2.50 and 2.63 children in their original families for male and female partners, respectively), but the frequency differences are not significant. Even pooling the sexes still yields an insignificant chi-square ($\chi^2 = 1.6$, $df = 1$, $p < .25$). Compatibility of partners by sibling positions itself may be considered the significant source of variation in this sample of marital behavior.

The test was repeated with all divorced couples (including those whose partners were single children) encountered among some 2,300 families studied in Nuremberg and Zurich under the sponsorship of the *Deutsch Forschungsgemeinschaft*. The 2,300 families provided the expected frequency values. Of 108 divorced couples whose data were complete, 18 showed high complementarity of their sibling positions (expected frequency 32), 52 couples displayed medium to low complementarity (expected frequency 49), and 38 couples displayed the lowest degree of complementarity (expected frequency 27; $\chi^2 = 10.8$, $df = 2$, $p < .01$). In other words, low complementarity of sibling positions of partners is more frequent and high complementarity is less frequent among divorced couples than in the population of marriages at large.

Even among divorced couples themselves, those who held out longer before the divorce occurred should have greater frequencies of high complementarity of sibling positions than those couples who were divorced early in their marriage. This is what the duplication theorem would lead one to expect. It predicts more permanence for those relationships that resemble early relationships more closely. Dividing the divorced couples that had provided adequate data into those who had been married for less than 10 years and those who had been married for 10 years or more before they were divorced yielded the frequency distributions for different degrees of complementarity of sibling positions shown in Table 2.

TABLE 2

DURATIONS OF MARRIAGE AMONG DIVORCED PARENTS IN RELATION TO COMPLEMENTARITY OF THEIR SIBLING POSITIONS

Sibling positions of parents	No. of divorced couples married for less than 10 years	No. of divorced couples married for 10 years or more
High complementarity	2 (7)	13 (8)
Medium to low complementarity	22 (20)	20 (23)
Lowest complementarity	17 (14)	14 (16)

Note.—Expected frequencies are in parentheses. $\chi^2 = 8.2$, $df = 2$, $p < .05$; when corrected for continuity: $\chi^2 = 6.5$, $df = 2$, $p < .05$.

As can be seen, the prediction is confirmed. Complementarity of sibling position is a factor in determining how long couples stay together. The higher the complementarity of their sibling positions, the greater the likelihood of postponement of ultimate divorce.

Of 108 divorced parents, 29 had had one (or more) children born out of wedlock before they married. The expected frequency based on the number of families in the sample of the population at large that had had at least one child born out of wedlock would be eight. The probability of such a frequency difference is very small ($p < .001$). It appears that those parents whose marriages end in divorce are also more likely to have been trapped into their marriage to begin with by the arrival of children. Had it not been for those children, many of them may never have married. Note that children "on the way" at the time of marriage have not been included in these considerations.

At any rate, those 29 couples constitute a subsample of all divorces and the author could check with them again to see whether they show lower frequencies of high complementarity of sibling positions than the population of marriages at large. Among those 29 divorced couples who had had a child born out of wedlock before they married, 5 showed high complementarity (expected 9), 6 medium to low complentarity (expected 13), and 18 lowest complementarity (expected 7; $\chi^2 = 22.9$, $df = 2$, $p < .001$). In contrast, the distribution of those parents from intact marriages who had had at least one child before they were married over the three classes of complementarity of sibling position of partners did not differ significantly from chance.

Summarizing, it was found that divorced couples showed lower degrees of complementarity of their sibling positions than the population at large. Among divorced couples, those who divorced early showed lower degrees of complementarity of sibling position than those getting their divorce late. Those among the divorced couples who had had at least one child before they married showed lower degrees of complementarity than the population at large. Current long-range social relationships such as marriage do indeed tend to be less enduring and less successful, other things being equal, the less they resemble early (intrafamilial, in particular sibling) relationships for the partners involved. They end in divorce, do so early, and may have been precipitated in the first place by the unintended birth of a child.

Indirect evidence suggests that more impressive results could probably have been obtained if those divorces that had not led to any children could also have been included in this study. Since all 2,300 couples studied were selected via schools, that is, their children, by definition there were no childless couples in the sample.

Divorced couples with partners who had been single children were included in this study. Hence it is more obvious than in the previous comparison of 16 divorced and 16 married couples that number of siblings of partners may itself be among the determinants of marital success. In order to test whether number of siblings of divorced partners differed from number of siblings of married partners in the population at large, the mean number of siblings of the partners without regard to the partners' sexes were computed for the three classes of complementarity of sibling positions. Table 3 shows the results.

TABLE 3

AVERAGE NUMBERS OF SIBLINGS OF MARRIAGE PARTNERS IN RELATION TO COMPLEMENTARITY OF THEIR SIBLING POSITIONS

Sibling positions of partners	Partners from intact marriages[a]	Partners from divorced marriages[b]
High complementarity	4.53	4.08
Medium to low complementarity	2.54	2.08
Lowest complementarity	2.54	2.30

Note.—In marriages of lowest complementarity of sibling positions (at least one partner being a single child), the average number of siblings refers to the nonsingle child among the partners.

[a] $N = 2{,}174$ (4,348 partners).

[b] $N = 108$ (216 partners).

As can be seen, the mean number of siblings of the partners to divorced marriages are consistently smaller than the mean numbers of siblings of married partners at large. In order to test for significance, classes of no sibling, one to three siblings, and four or more siblings were formed and the frequencies of divorced partners in each class compared with the frequencies of partners taken from the entire sample of the population (2,300 families). Table 4 shows the results.

It is apparent that divorced partners do come from somewhat smaller sibling configurations than would be expected by chance. It has already been suggested that larger numbers of siblings make it more likely for a partner to have a sibling relationship among them that is complementary to that of a given potential spouse. In order to test what part sibling role complementarity plays regardless of the number of the partners' siblings, the latter should be held constant. This was accomplished by selecting those divorced couples whose partners had each had one sibling only, plus those couples one of whose partners had had one and the other two siblings.

It was found that of 31 couples thus selected, 5 showed high complementarity (expected frequency 11) and 26 of them medium to low complementarity (expected frequency 20; $\chi^2 = 5.1$, $df = 1$, $p < .05$). Obviously, complementarity of sibling positions of partners is effective regardless of number of

TABLE 4

NUMBER OF SIBLINGS OF DIVORCED PARTNERS

Partner	No. of siblings		
	No sibling	1 to 3 siblings	4 or more siblings
Male	24 (15)	62 (61)	22 (32)
Female	23 (16)	66 (60)	19 (32)

Note.—Expected frequencies are in parentheses. Male partners: $\chi^2 = 8.5$, $df = 2$, $p < .05$; female partners: $\chi^2 = 9.0$, $df = 2$, $p < .05$.

siblings of partners. Marriage partners who had both come from two-children families or from a two-children and a three-children family displayed the same trend as did all 108 divorced couples: a lower incidence of high complementarity and a higher incidence of medium to low complementarity of sibling positions than marriage partners of the population at large.

Another indicator of the success of a relationship such as marriage would be children. They come about as a resultant of many determinants. Besides ethnic, religious, socioeconomic, political, geographic, and educational factors that influence the decisions of couples to have, or at least to risk the prospect of, children, we would add size of the partners' original families and the relative ages of the partners' parents. The duplication theorem would suggest that partners create families in accordance with their early experiences of family life and family relations. There is statistical evidence to this effect that has been reported elsewhere (Toman et al., 1967).

But, other things being equal, we might still argue that more children mean more success of the marriage relationship. Unhappy parents would tend to have fewer children or none, or to wish they had none and decide at least to have no more. Happy parents, in turn, would tend to be able to cope with their children better and to be inclined to have more of the same.

If this were indeed so, as seems likely by common sense and ample though anecdotal clinical evidence, let us check again whether complementarity of sibling position according to the duplication theorem—supposedly one of a number of prerequisites for marital success—is reflected, among other things, in the number of children that parents have. Do parents whose sibling positions are of high complementarity have more children than parents whose sibling positions are of low complementarity? Table 5 gives the answer.

TABLE 5

NUMBER OF CHILDREN IN FAMILY AS RELATED TO COMPLEMENTARITY OF THE PARENTS' SIBLING POSITIONS

Sibling positions of parents	No. of children		
	1	2 to 4	5 and more
High complementarity	110 (146)	501 (482)	71 (53)
Medium to low complementarity	205 (222)	749 (732)	80 (81)
Lowest complementarity	182 (129)	388 (424)	30 (47)

Note.—Expected frequencies are in parentheses. $\chi^2 = 48.4$, $df = 4$, $p < .001$.

As can be seen, parents of high complementarity of sibling position had indeed significantly more children than expected by chance, whereas parents one or both of whom had been single children (i.e., parents of the lowest degree of complementarity) had significantly fewer children. Broadly speaking, high complementarity of sibling positions of parents, on the average, makes for "half a child" more than does lowest complementarity. Since the average number of children in the families studied was approximately 2.7, this quantity of "half a child" is not negligible.

Marriages at an early age could be expected to be based on less experience with potential partners than marriages at a later age. One could assume that persons selecting a partner at a later age have had potential access to all those partners whom they might have selected at an earlier age, plus those whom they have come to know since then. Hence if people go by the duplication theorem at all, wittingly or unwittingly, when selecting permanent mates, those with more experience should heed it better than those with little experience. In other words, people marrying relatively late should be more likely to choose partners of complementary sibling positions than would be those marrying relatively early. Dividing partners into those married below the median age of marriage and those married at or above the median age, we obtain Table 6.

As can be gathered from Table 6, male and female partners alike tended to choose partners of complementary sibling position more often than chance when they had married late. They tended to choose such partners less frequently than chance when they had married early.

One component of the decision of parents to have another child may be the configuration of children that they have had thus far.

TABLE 6

AGES AT MARRIAGE IN RELATION TO COMPLEMENTARITY OF SIBLING POSITIONS

Sibling positions of parents	Husband's age at marriage[a]		Wife's age at marriage[b]	
	Less than 27 years	27 years or more	Less than 24 years	24 years or more
High complementarity	275 (311)	387 (351)	281 (322)	381 (340)
Medium to low complementarity	453 (454)	513 (512)	459 (470)	507 (496)
Lowest complementarity	288 (251)	248 (285)	313 (261)	223 (275)

Note.—Expected frequencies are in parentheses.
[a] $\chi^2 = 18.1$, $df = 2$, $p < .001$.
[b] $\chi^2 = 30.9$, $df = 2$, $p < .001$.

This may be so regardless of complementarity or noncomplementarity of sibling positions of parents, and may tend to confound the efficacy of the duplication theorem.

As it turned out, parents did indeed respond differentially to the sexes of their children born to them thus far. If they had boys first or boys only, they were more apt to have still another child than were those who had girls first or girls only (see Table 7). If their first child was a boy, b, they had, on the average, 1.51 more children. If their first child was a girl, s, they had, on the average, 1.38 more children. If their first two children were boys, bb, they had 1.07 more children; if their first two children were girls, ss, they had merely .71 more children. If their first three children were boys, bbb, they had still 1.06 more children; if their first three children were girls, sss, they had only .50 more children, etc.

If one assumes that parents try for more of the same, if they like what they got so far, one could argue that they wanted boys a little more than they did girls. This would not preclude that they may not have been delighted to have a girl after a series of boys, but trying for another child they would, of course, risk getting still another boy. They seemed more willing to take that risk than did the parents of a succession of girls who, in turn, may have been equally or even more delighted to get a boy.

One could argue otherwise: Parents are more satisfied by a succession of girls and need no more children, particularly since it could also be boys.

Be that as it may, the differential effect that particular configurations of children have on the decisions of their parents to have more children or not has been demonstrated.

Some such argument could perhaps be made in favor of all the data and findings reported here. The reader may have little trouble accepting the incidence of divorce or the short duration of marriage before divorce occurs as some indication of failure of the relationship. He may be less apt to accept number of children born to a couple as

TABLE 7

NUMBER OF SUBSEQUENT CHILDREN IN THE FAMILY IN RELATION TO DIFFERENT SEX DISTRIBUTIONS AMONG CHILDREN BORN THUS FAR

Children born thus far	Frequencies of families						Total number of families	Average number of subsequent children
	No. of subsequent children							
	0	1	2	3	4	5 etc.		
b	72	108	59	38	19	9	305	1.51
s	73	117	67	20	16	8	301	1.38
bb	54	29	16	18	2	1	120	1.07
bs	54	28	22	3	6	1	114	.96
sb	54	36	9	10	2	1	112	.87
ss	63	34	10	5	2	1	115	.71
bbb	17	9	6	1	1	2	36	1.06
sss	15	1	3	1	0	0	20	.50
bss	17	14	1	5	1	0	38	.92
sbb	13	6	6	2	0	1	28	1.04
bbs + bsb	29	19	20	1	1	0	70	.94
ssb + sbs	47	12	6	5	2	1	73	.65

Note.—To test for significance by χ^2, different frequency classes had to be combined wherever frequency entries had been too low by themselves. Hence the varying degrees of freedom: All b vs. all s: $\chi^2 = 6.87$, $df = 4$, *ns*; all bb vs. all ss: $\chi^2 = 8.16$, $df = 3$, $p < .05$; all bbb vs. all sss: $\chi^2 = 6.28$, $df = 2$, $p < .05$; all bbs + bsb vs. all ssb + sbs: $\chi^2 = 8.30$, $df = 2$, $p < .02$.

an indicator of relative marital success, and above-average age of partners at the time of marriage as an indicator of their greater maturity when choosing a spouse. What can hardly be denied are the deviations of frequency distributions from chance that were observed and reported and that had not been known before. One may say in favor of the duplication theorem that it is at least at no great loss to explain those various deviations from chance in a somewhat consistent manner.

To show the scope of this consistency, it should be mentioned that the duplication theorem has been shown to apply also to friendships among persons of the same sex. Those friends who were of complementary sibling-age-ranks developed longer lasting friendships, particularly when one or both of them had come from monosexual sibling configurations, than did those friends who were of identical or similar sibling-age-rank (see Toman, 1969).

The duplication theorem applies also to the parent–child relationship. It could be shown that the parent and same-sex child identify more readily with one another if they hold similar or identical sibling positions. Parent and opposite-sex child, on the other hand, interact better with one another if their sibling positions are complementary to each other (Toman, 1962a, 1965, 1969).

Finally, the duplication theorem has been found to apply also to losses of persons suffered in the original family. They create expectations of further loss and, as a consequence, poorer, more haphazard, choices of partners, friends, and spouses, more conflict-laden relationships (in the present case defined in terms of incompatibility of sibling positions of partners) and greater likelihoods of new losses than expected by chance (see Toman, 1962a, 1965, 1969). Among the two samples of the normal population (taken from Nuremberg and Zurich), it was found that parents of both sexes who had suffered the loss of a parent themselves took significantly longer than expected by chance to have children of their own. More often than parents without such losses, they were either unmarried or divorced. This was significant for fathers regardless of whether they had lost their father or their mother, for mothers only when they had lost their fathers (Toman et al., 1967).

DISCUSSION

These effects of family constellation characteristics (and other effects reported elsewhere) are only one among several groups of effects on long-range social behavior. They are not overwhelmingly strong themselves, but wherever family constellation characteristics known to be effective and independent of each other occur in combinations, their influence becomes more pervasive. The fact that they could be demonstrated at all may be considered a surprise in view of the fact that these effects extend over time spans of some 20 years.

These effects do not bridge the time spans by magic, to be sure. The sibling roles learned in the original family, first at a time when the child experiences social contacts other than within the family only rarely and briefly, may be applied in new social contexts such as on playgrounds, around church, in school and clubs, etc. By the time those other contexts begin to provide more substantial and extensive opportunities for experience, patterns for evaluating those opportunities have already been set. These patterns are capable of modification and correction, but any such changes will be harder to accomplish, the longer and more exclusively those patterns have already prevailed. Individuals learn to generalize from family members to playmates and classmates, colleagues, acquaintances, friends, and ultimately those friends whom they decide to keep or marry. If these generalizations happen to work in new contexts, the patterns of evaluation and social behavior are being confirmed. If they lead to negative experiences repeatedly or even once, but drastically or very unexpectedly, social avoidance behavior will result and perpetuate itself, since avoidance precludes further and possibly corrective experiences.

One may argue that the duplication theorem would plausibly hold where family circumstances were favorable. Under unfavorable circumstances, especially when due to conflicts among family members and losses of family members, the individual should fare better when searching for other or even opposite opportunities for permanent social relationships than those encountered in his family. Persistent anecdotal evidence and some statistical data suggest that even individuals coming from unfavorable circumstances tend to do better and to be more successful in new social contacts when they do not turn to the opposite of what they had at home. Mild changes for the better appear to be more trustworthy than radical changes for the better.

Socioeconomic strata also exert influences on permanent social relationships. Families of the lower classes tend to marry sooner, to come from larger families, and to have more children, but also more frequently to suffer losses such as absences or deaths of fathers, separations from mothers, losses of siblings, etc. This includes also temporary losses (absences of family members for months or years until they eventually reappear) and partial losses (discoveries, e.g., that father has been in jail, that he is a drunkard, that mother never went to high school, or that she has had and given away a child before she was married, etc.).

The sample of the population studied can be called urban middle class including small percentages of lower-middle and of upper-middle class. The samples of the schools from which the parents were taken had been matched by type, size, and relative location. There is no denying, though, that we have studied only Western European urban industrial populations. In rural areas, in other geographic regions, under different ethnic or political conditions, some of the influences we have traced may be contaminated by other determinants. Studies of smaller samples of urban populations in the United States, however, showed very similar results to those reported here.

With increasing size of a sibling configuration, there are more middle siblings and hence more multiple sibling roles. More pronounced roles (oldest, youngest, only boy, only girl, etc.) are in the minority. Consequently, all sibling roles may appear somewhat blurred in those larger sibling configurations. With the overall trend of our times toward urbanization and smaller families, however, sibling roles may tend to become generally more conspicuous and the present findings relevant for increasing numbers of family constellations.

The study presented is in the tradition of the work of Bowlby (1951), Glueck and Glueck (1959), Hollingshead and Redlich (1958), Strotzka (1965), and others active in the area of the social psychology of long-range interpersonal relationships and of social psychiatry. The data and findings are of a type that may have escaped the notice of factor-analytic personality research (e.g., Brill & Beebe 1955; Cattell 1964; Eysenck 1961), even though this multivariate technique of analysis requires as much inclusiveness as possible in its respective primary data.

REFERENCES

ADLER, A. *Praxis und Theorie der Individualpsychologie.* München: Bergmann, 1920.

BOWLBY, J. *Maternal care and mental health.* London: World Health Organization Monograph, 1951.

BRILL, N. Q., & BEEBE, G. W. *A follow-up study of war neuroses.* Washington, D. C.: Veterans, Administration Medical Monograph, 1955.

CATTELL, R. B. *Personality and social psychology (collected papers).* San Diego, Calif.: Knapp, 1964.

EYSENCK, H. J. The effects of psychotherapy. In H. J. Eysenck (Ed.), *Handbook of abnormal psychology, an experimental approach.* New York: Basic Books, 1961.

FREUD, S. *A general introduction to psychoanalysis (1916/17).* Vol. 15 and 16. (Standard edition of the complete works of Sigmund Freud) London: Hogarth Press, 1953.

FREUD, S. *Beyond the pleasure principle (1920).* Vol. 18. (Standard edition to the complete works of Sigmund Freud) London: Hogarth Press, 1953.

GLUECK, S., & GLUECK, E. *Predicting delinquency and crime.* Cambridge, Mass.: Harvard University Press, 1959.

HOLLINGSHEAD, A. B., & REDLICH, F. C. *Social class and mental illness.* New York: Wiley, 1958.

HULL, C. L. *Principles of behavior.* New York: Appleton-Century-Crofts, 1943.

KOCH, H. L. Some personality correlates of sex, sibling position and sex of sibling among five-

and six-year-old children. *Genetic Psychology Monographs,* 1955, **52**, 3–50.

KOCH, H. L. Children's attitudes toward their peers as related to certain characteristics of their siblings. *Psychological Monographs,* 1956, **70** (19, Whole No. 426).

KOCH, H. L. The relation of certain formal attributes of siblings to attitudes held toward each other and toward their parents. *Monographs of the Society for Research in Child Development,* 1960, **25**(4), No. 78.

KROUT, M. H. Typical behavior patterns in 26 ordinal positions. *Journal of Genetic Psychology,* 1939, **55**, 3–30.

SCHACHTER, S. Birth order and sociometric choice. *Journal of Abnormal and Social Psychology,* 1964, **68**, 453–456.

SCHOONOVER, S. M. The relationships of intelligence and achievement to birth order, sex of sibling, and age interval. *Journal of Educational Psychology,* 1959, **50**, 143–146.

STROTZKA, H. *Einführung in die Sozialpsychiatrie.* Hamburg: Rowohlt, 1965.

TOMAN, W. Die Familienkonstellation und ihre psychologische Bedeutung. *Psychologische Rundschau,* 1959, **10**, 1–15. (a)

TOMAN, W. Family constellation as a basic personality determinant. *Journal of Individual Psychology,* 1959, **15**, 199–211. (b)

TOMAN, W. Family constellation as a character and marriage determinant. *International Journal of Psychoanalysis,* 1959, **40**, 316–319. (c)

TOMAN, W. Family constellation. New York: Springer, 1962. (a)

TOMAN, W. Family constellations of divorced and married couples. *Journal of Individual Psychology,* 1962, **18**, 48–51. (b)

TOMAN, W. *Familienkonstellationen.* München: Beck, 1965.

TOMAN, W. *Family constellation.* (Rev. ed.) New York: Springer, 1969.

TOMAN, W., PREISER, S., GASCH, B., & PLATTIG, G. *Die Wirkungen von Familienkonstellationen auf Personen, ihre sozialen Beziehungen und die nachfolgende Generation.* (Projekt der deutschen Forschungsgemeinschaft 1962–1966.) Bad Godesberg, 1967.

WARREN, J. R. Birth order and social behavior. *Psychological Bulletin,* 1966, **65**, 38–49.

(Received April 13, 1970; revision received March 15, 1971)

Psychological Review
1971, Vol. 78, No. 5, 391–408

COMPONENTS OF ATTENTION[1]

MICHAEL I. POSNER[2] AND STEPHEN J. BOIES

University of Oregon

The study of human attention may be divided into three components. These are alertness, selectivity, and processing capacity. This paper outlines experimental techniques designed to separate these components and examine their interrelations within comparable tasks. It is shown that a stimulus may be used to increase alertness for processing all external information, to improve selection of particular stimuli, or to do both simultaneously. Development of alertness and selectivity are separable, but they may go on together without interference. Moreover, encoding a stimulus may proceed without producing interference with other signals. Thus, the contact between an external stimulus and its representation in memory does not appear to require processing capacity. Limited capacity results are obtained when mental operations such as response selection or rehearsal must be performed on the encoded information.

The subject of attention is a broad one. There are many definitions of it and subcategories included under it. In a recent book, Moray (1970) has proposed at least six different meanings of the term attention in current psychological research. It would be premature to propose any single taxonomy or exhaustive set of categories for the term. However, both in looking at the various definitions that Moray proposes and in reviewing the literature, there do appear to be three major topics under which studies of attention might be grouped.

First is the notion of alertness. Maintaining attention in the sense of alertness[3] is presumably involved in human ability to perform in long, boring tasks like those that psychologists design to study vigilance (Mackworth, 1970). The ability to develop and maintain an optimal sensitivity to external stimulation is also studied when subjects (*S*s) receive a warning to prepare themselves to take in information. The foreperiod of a reaction time task may be considered as a miniature vigilance situation where alertness must be developed rapidly and maintained over a relatively brief interval. Recent studies have investigated the time course (Bertelson, 1967) and the brain activity (Karlin, 1970; Näätänen, 1970; Posner & Wilkinson, 1969) during this interval. It appears that there is some similarity between the brain processes which take place during the foreperiod and those involved in vigilance tasks (Wilkinson & Haines, 1970). These studies provide some rationale for including alertness as a general component of attention.

A second sense of attention is the ability to select information from one source or of one kind rather than another (Egeth, 1967; Triesman, 1969). Studies of selective attention may require that *S*s report information from a particular sensory modality, spatial location, or content (letters rather than digits). Broadbent (1958) has argued that at least some aspects of selection are performed by filtering mechanisms which block out or attenuate input. Other theories have stressed additional operations which are performed on information of particular importance (Deutsch & Deutsch, 1963; Norman, 1968). These views bear a relationship to physiological models of attention that have stressed either a tuning process which blocks input from unselected sources (Hernandez-Peón, 1966) or a general alerting reaction which enhances input from whatever source (Näätänen, 1970). The former view is more in line with filtering mechanisms, while the latter fits better with the idea of additional operations performed on highly processed input. One way of viewing this issue is to ask whether *S*s can prepare for signals in a way which is not selective. If they prepare for a visual signal, is the capacity to process auditory signals decreased or enhanced over an unprepared state? Thus, the separability of alertness and selection as internal processes has important consequences for a theory of attention.

---

[1] Portions of this paper were presented in an invited address to the Western Psychological Association, Los Angeles, April 1970. The research included here was sponsored by National Science Foundation Grants GB 5960 and 21020 and by the Advanced Research Projects Agency of the Department of Defense monitored by the Air Force Office of Scientific Research under Contract No. F44620-67-C-0099 all to the University of Oregon. The authors are grateful to Barbara Lewis and Larry Moore for help in running the subjects.

[2] Requests for reprints should be sent to Michael I. Posner, Department of Psychology, University of Oregon, Eugene, Oregon 97403.

[3] The term alertness is used here rather than arousal to avoid the emotional connotation of many uses of the latter term. The degree of separation of the terms is not clear at this point.

A third sense of attention relates to the idea of a limited central processing capacity (Broadbent, 1958). It is well known that *S*s have difficulty in handling two tasks simultaneously, even when they wish to do so. Signals which arrive during the reaction time (RT) period tend to be delayed (Welford, 1968). Yet it is not clear what aspects of signal processing are subject to the single-channel limitation. Some have argued that the single channel extends to almost all processing (Welford, 1968), while other have seen it as applying mainly to response execution (Reynolds, 1964). A limited capacity mechanism suggests that any two operations requiring it will interfere with each other. Thus, a study of the ability of *S*s to time share mental operations can lead to an understanding of what aspects of processing require access to the limited capacity system.

Most recent articles related to attention involve one of these three components (see e.g., Koster, 1969; Sanders, 1967, 1970). Unfortunately, the experiments used to study each of the components are different and there has been little attempt to develop tasks which allow each of them to be separated and interrelated. Moray (1970) suggests the importance of doing so as follows:

> It might well be, for example, that the relation between selective listening and arousal is such that arousal level acts as a parameter which will alter the over-all efficiency of selection and rejection as it varies. But, no systematic investigation has so far been carried out [p. 85].

This paper is an attempt to develop means for studying the separation and interrelationship of these components of attention at the behavioral level.

Consider first a task which involves all three components mixed in a fairly complex way. The *S*s see a single letter for a brief period of time, followed after a variable interval by a second letter. The task is to respond as rapidly as possible "same" if the two letters have the same name; otherwise, "different" (Posner, 1969; Posner, Boies, Eichelman, & Taylor, 1969). The first letter has two functions. It serves as a warning signal and also tells the *S* what letter he is looking for. Thus, one might expect it to increase alertness and to introduce some bias of the perceptual system toward tests related to the configuration of the first letter. The bias might involve, for example, activation of trace systems related to the physical configuration of the letter (Posner, 1969). What happens during the interval between the two letters depends greatly on how the attention (processing capacity) of *S* is directed. For example, if he knows that all matches will be physically identical, he maintains his speed on physical matches much better than if some matches must be based on the name (Posner, 1969). If the first letter disappears, there is a rapid increase in times for a physical match, while name matches remain relatively constant, suggesting that *S*s keep the letter name in mind. However, if the physical form remains present during the interval, physical matching remains constant while name matches increase in efficiency until they resemble the physical matches. This suggests that *S*s generate a visual code of the opposite case (Boies, 1969; Posner, 1969). These results indicate that *S*s are very flexible in the coding they use for letters and similar visual material (Cohen, 1969; Tversky, 1969), but their choices are constrained by the rivalry between physical and name codes for *S*s' limited processing capacity.

This brief description of the letter-matching sequence suggests that it may be a useful task for separating the three components of attention described above. The experiments reviewed in this paper are an effort to do so. Moreover, they attempt to connect the operations of this experimental paradigm to the three theoretical components of attention outlined above. Two major questions are considered. First, Can the warning function and the selective function of the first letter be separated? If so, how do the two functions relate? Second, What is the relationship between these two components of attention and central processing capacity?

PREPARATION

Many investigators have studied the period of time between a warning signal and a signal related to the required response (Bertelson, 1967). Of primary interest are studies where *S* has little uncertainty about when the response-related signal will occur. This is achieved by using blocks of trials with a fixed warning interval. Whether the response task involves reaction time (Bertelson, 1967), signal detection (Egan, Greenberg, & Schulman, 1961), or other types of processing (Leavitt, 1969), the results are similar. Performance is worse with no warning and improves as the warning interval increases to some optimal value. This value is most frequently about .2–.5 second. An interval longer than optimal tends to produce a decrease in performance, but this is usually not as pronounced, particularly when feedback is used to keep motivation high.

Figure 1 indicates functions obtained using two-letter matching tasks (Posner & Wilkinson, 1969). Both tasks involved a warning signal followed after a foreperiod by a pair of letters. The foreperiod was constant for a block of trials but varied between blocks. The *S*s had a single key which they pressed when the letter pair was defined as "same" for that task. The solid line is for a task which required *S* to press a key if the letters were physically identical, while the dotted line required him to press the key if both letters were either vowels or consonants. Although RT differs greatly in the two tasks, the time course of preparation is very similar and like that described in the studies cited above.

ENCODING

If *S* is provided with part of the information he needs to perform a speeded task, his RT when he receives the remainder will be reduced (Cohen, 1969; La Berge, Van Gelder, & Yellott, 1970; Leonard, 1958). If the time between the two signals is varied, it is possible to discover how long it takes to encode the first signal in a form which is op-

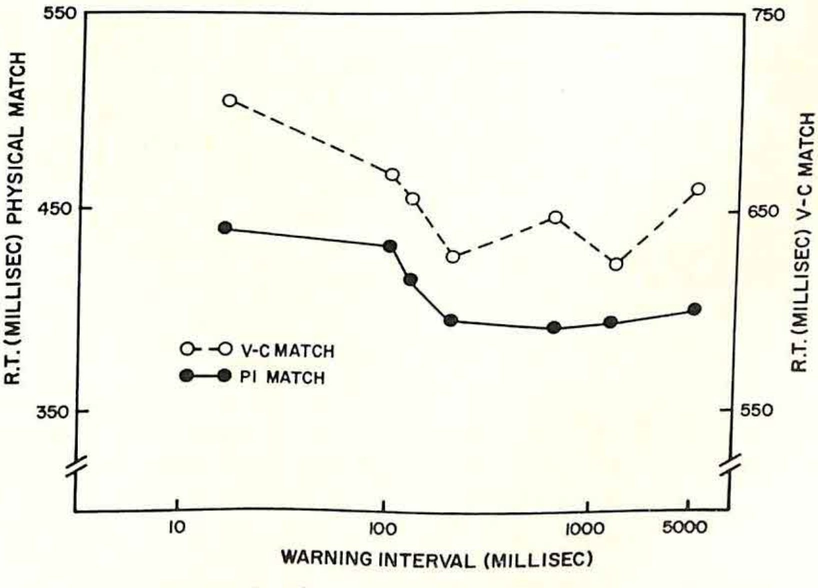


FIG. 1. Preparation functions for physical (PI) match task (solid line, left ordinate) and vowel-consonant (V–C) match task (dotted line, right ordinate). (Data are from constant block foreperiods of each warning interval collapsed across four days; after Posner & Wilkinson, 1969.)

timal for processing the second. If the first signal is a letter that must be matched against a second letter, the improvement should represent the encoding of the information from the first letter needed for the match. In order to reduce the involvement of general preparation in such functions, a warning signal was given .5 second before the first letter. It was supposed that preparation could be maintained during the time between presentation of the first letter and the presentation of the second letter.

Method

Experiments were run on a PDP-9 computer. Displays were presented on a Tektronix 503 scope using a P 11 phosphor which decays to .01 of its energy in 20 milliseconds (msec.). The warning signal was a cross presented .5 second before the first letter. The first letter was presented directly below the warning signal and remained on until *S*'s response. After a variable interval (interstimulus interval [ISI]), the second letter was presented adjacent to the first. At a distance of 16 inches the two letters were foveal (3 degrees of visual angle). The time between the two letters was varied between blocks of trials.

Two separate experiments were run. In the first experiment, *S* responded "same" if the letters were physically identical; in the second, "same" was defined as both vowels or both consonants. Experiment I used only capital letters, while Experiment II used mixed uppercase and lowercase letters. The computer chose the letters at random from an alphabet of 15 letters with the restriction that half of the trials in each block be "same." In these experiments, *S*s had two keys and were instructed to press the left one if the letters were defined as "same" and the right one if "different."

In Experiment I, each of eight *S*s ran in six 40-trial blocks for each of four days. Each block contained one of six intervals between the two letters. The intervals were assigned randomly so that each *S* had all six blocks on each day. Experiment II was substantially the same except that each block contained four physical matches (e.g., AA), eight name matches (e.g., Aa), 12 vowel consonant "same" responses, and 24 "different" responses.

Results

First, it is clear from Figures 2 and 3, as found previously (Posner & Mitchell, 1967), that the level of match has systematic effects on the reaction time. Moreover, the encoding function for all three

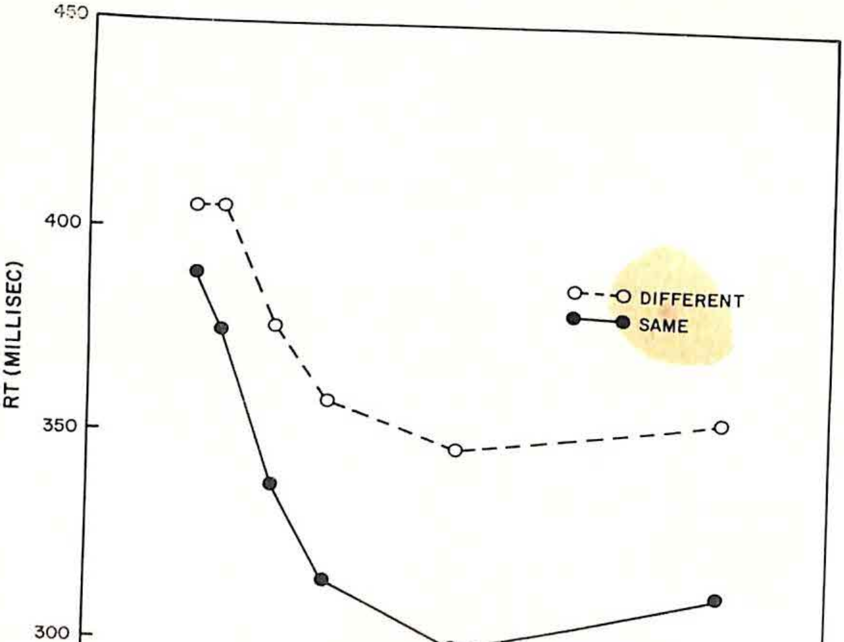


FIG. 2. Encoding function for physical matches. (Constant warning interval of 500 msec., with constant blocks of various ISIs.)

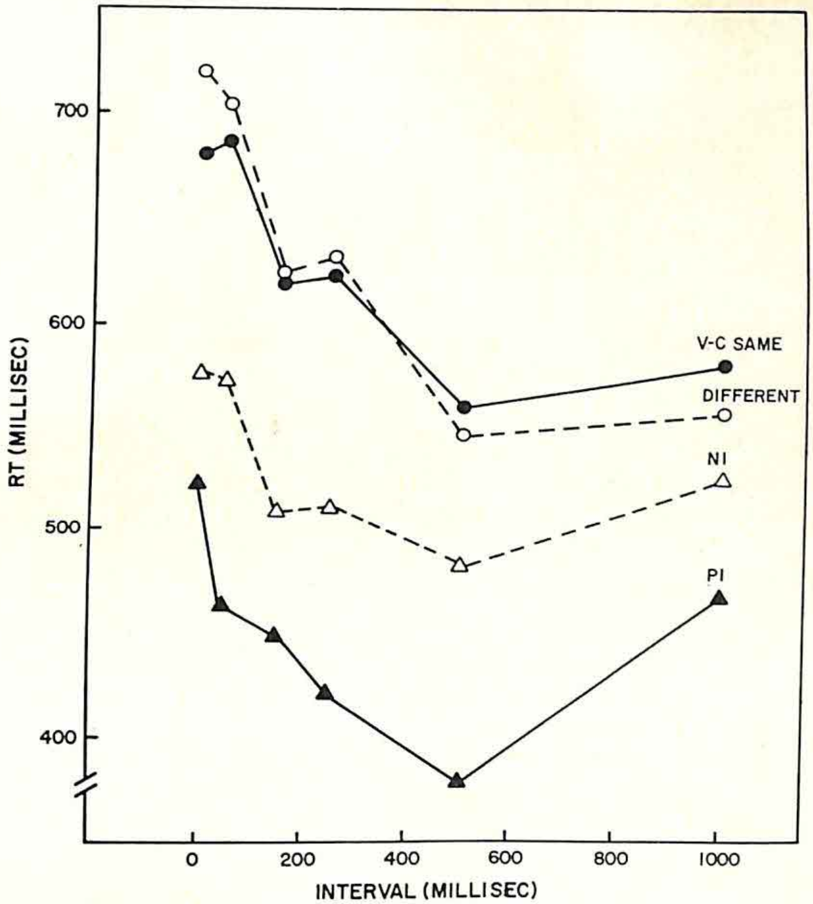


FIG. 3. Encoding functions for vowel–consonant (V–C) match task. (Constant warning interval of 500 msec. with constant blocks of various ISIs. NI = name identity, PI = physical identity.)

levels is strikingly similar in both experiments. The optimal RT is at 500 msec. in all functions. This finding appears to be robust over the distribution of intervals, since a control study using intervals of 0, 1, 50, 150, 250, and 500 showed almost the same function as that obtained for the first five points of Figure 2. The same result is also obtained in encoding functions discussed in the next section (see Figures 4 and 5). There is some tendency for sharpest improvement in RT to occur later for more complex matches. Most of the improvement for physical matches appears to come in the first 150 msec. This is much less true of vowel–consonant matches, which show a sharp improvement between 250 and 500 msec. The extent of improvement appears to be slightly greater for vowel–consonant matches than for other types, but name matches do not show much more improvement than physical matches.

In general, encoding and preparation functions appear to be rather similar. Both show an optimal point at about 500 msec. Both tend to give greater improvement for vowel–consonant than for physical matches. The next experiment is designed to compare them more directly.

PREPARATION AND ENCODING

If *S* is given 500 msec. to prepare, his RT will be about optimal. If he is given information at the end of that 500 msec. about what he is to select, he shows a further improvement which also goes on for about 500 msec. Are these two improvements due to two separate processes, or are they merely two different ways of studying the same overall process? A paradigm was designed

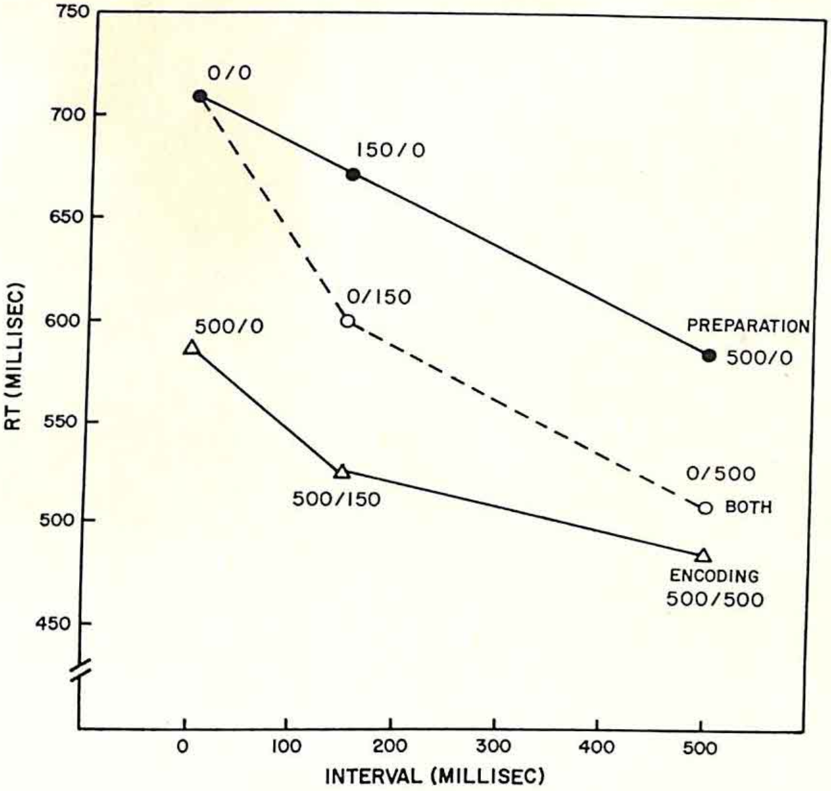


FIG. 5. Preparation, encoding, and "both" functions for name matching task. (Values of WI/ISI are shown for each point.)

Results

Three functions can be drawn from the seven conditions. Each condition involving a combination of one WI and one ISI. A preparation function involves conditions 0/0, 150/0, and 500/0 (WI/ISI). This function is shown in the upper curve of Figures 4, 5, and 6.

The encoding function is obtained from conditions where *S* has always had a full 500-msec. WI and either 0-, 150-, or 500-msec. ISI. This is shown in the lower curve of each figure. When there is no warning signal, the first letter serves both as a warning signal and to provide selective information. The function obtained from conditions 0/0, 0/150, and 0/500 is called the "both" function and is shown as the middle line of each figure. The figures show the mean of the median RTs for each condition.

The main effects of WI and ISI on RT are highly significant in all experiments. In all three experiments there is an interaction between WI and ISI. An interaction between two independent variables has been interpreted (Sternberg, 1969) as indicating that they affect the same stage of processing. In the present case, WI and ISI may be thought to interact because they both affect alertness. At zero WI, alertness must be varying during the ISI, while with an optimal WI, *S* may already be alert and thus ISI would mainly involve encoding of information from the first letter.

Table 2 shows the amount of improvement (from the 0/0 condition for preparation and "both" and from the 500/0 for encoding) due to preparation, encoding, and both together. The results clearly indicate that in the case of physical and name instructions, the improvement obtained in the "both" condition is almost exactly the sum of that obtained separately from the preparation and encoding functions. In the vowel–consonant case, there appears to be a more substantial departure from additivity, but these are not significant due to high variability among *S*s in this condition.

TABLE 1

Conditions of Warning and Interstimulus Intervals for Experiments on Encoding and Preparation

Interstimulus interval[a]	Warning interval[a]		
	0	150	500
0	Preparation Both	Preparation	Preparation Encoding
150	Both		Encoding
500	Both		Encoding

[a] In milliseconds.

to compare them directly. The method used was to obtain preparation functions, encoding functions, and functions where both processes might go on simultaneously.

Method

Three experiments were run involving a common method. The apparatus and displays were the same as discussed previously. Three conditions varied only the warning interval (WI) and were used to produce a preparation function. Three conditions involved an optimal WI (.5 second) and varied only ISI. These produced an encoding function. Three conditions used zero foreperiod and varied ISI. In this case, the first letter served as both a warning signal and as a source of specific information. Two conditions enter into two functions, making a total of seven different combinations of WI and ISI as shown in Table 1. Each condition involved a block of 40 trials; 20 "same" and 20 "different." Each trial was separated by a long variable intertrial interval of 8–15 seconds. This was followed by a warning signal, the first letter and the second letter. The warning signal remained on only during the WI, while the first and second letters remained present until *S*s responded.

Each experiment involved 10 *S*s run in one block of each of the seven conditions in a random order on each of two days. The first day was considered practice. After each trial, *S*s were given feedback as to the correctness of their responses and the RTs. The three experiments differ only in the level of instruction. In the first experiment, "same" was defined as physical identity and all letters were uppercase; in the second experiment, "same" responses were defined as letters with the same name; while in the final experiment, "same" was defined by the rule both vowels or both consonants. In all experiments, the "same" response was made with the left index finger and the "different" response with the right index finger.

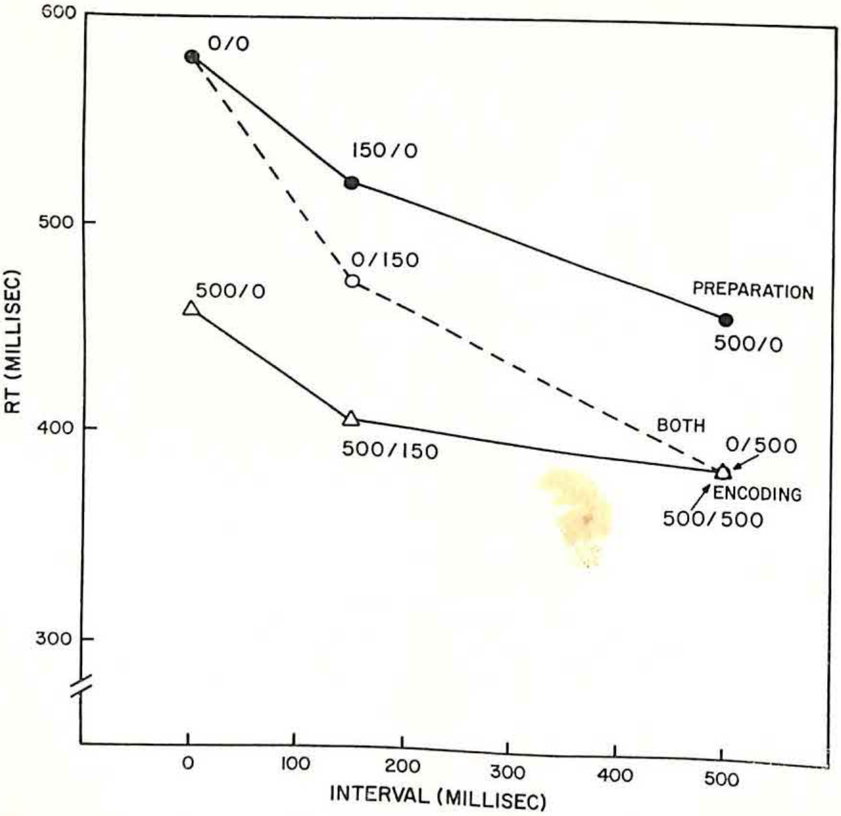


Fig. 4. Preparation, encoding, and "both" functions for physical matching task. (Values of WI/ISI are shown for each point.)

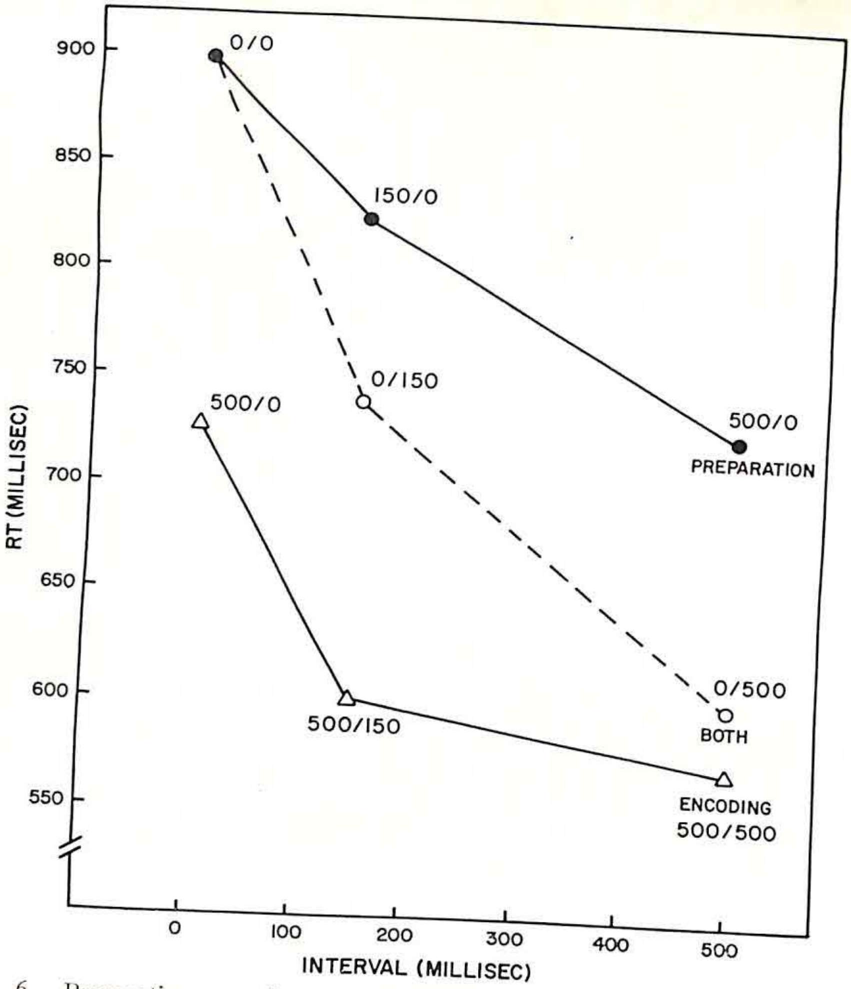


FIG. 6. Preparation, encoding, and "both" functions for vowel-consonant matching task. (Values of WI/ISI are shown for each point.)

The additivity of improvement due to preparation and encoding suggests that the two processes can proceed in parallel without interference. This is true even at 150 msec. when the functions are changing rapidly. Moreover, the additivity indicates that the amount of improvement due to encoding is the same regardless of the initial level of alertness.

It remains to be shown that the encoding function is really related to specific information from the letter presented. A striking finding in the physical match encoding function is a significant interaction between interval and type of response (same vs. different) $F = 3.9$, $df = 2/9$, $p < .05$. Recently a number of investigators have pointed out that the response "different" does not appear to be the mirror image of the response "same" (Bamber, 1969; Nickerson, 1965; Smith & Nielsen, 1970; Tversky, 1969). In many experiments where there is an interval between the two items to be matched, "same" responses are faster than "different" responses (Bamber, 1969; Nickerson, 1965). Figures 7 and 8 show the data from the preparation, encoding, and "both" functions for the physical match. It is clear that the relative efficiency in processing "same" as compared to "different" stimuli increases as a function of interval, but only when *S* has been given the first letter. The relative increase in efficiency of the "same" response as compared to "different" shown in the RTs is also reflected in the error rates. No such change is found when the interval involves the period between a warning signal and the letter pair. Nor does it occur for name or vowel-consonant level matches.

The finding that "same" and "different"

responses are identical in simultaneous presentation has not been universal (Krueger, 1970; Posner & Mitchell, 1967). This appears to depend on many factors. It was suspected that one such factor was assignment of hands. To check on this, a control study was run which assigned six *S*s "same" on the left and six *S*s "same" on the right. The experiment involved ISIs of 0, 50, 150, and 500 msec. with a 500-msec. WI. The physical match instruction was used. The data for left-hand "same" *S*s replicated Figure 7 perfectly. For *S*s assigned "same" on the right, however, "same" was faster even with simultaneous presentation. However, regardless of hand assignment, the significant divergence over time between "same" and "different" responses was found (see also Figure 2 and Corballis, Lieberman, & Bindra, 1968).

The finding that presentation of the first letter improves the efficiency of handling an identical letter suggests that the first letter changed *S*'s sensitivity to a letter of identical form. One way of conceptualizing this is in terms of Sokolov's neural model idea (Sokolov, 1963). The first letter serves as

TABLE 2

AMOUNT OF IMPROVEMENT IN REACTION TIME FROM PREPARATION, ENCODING, AND BOTH

Match	Interval[a]	
	150	500
Physical		
Preparation	58	122
Encoding	53	73
Both	101	195
Name		
Preparation	42	127
Encoding	58	98
Both	111	201
Vowel–Consonant		
Preparation	71	170
Encoding	124	150
Both	157	296

Note.—Improvement is measured by subtracting RT at specified interval from RT at 0/0 for preparation and "both" function and 500/0 for encoding.

[a] In milliseconds.

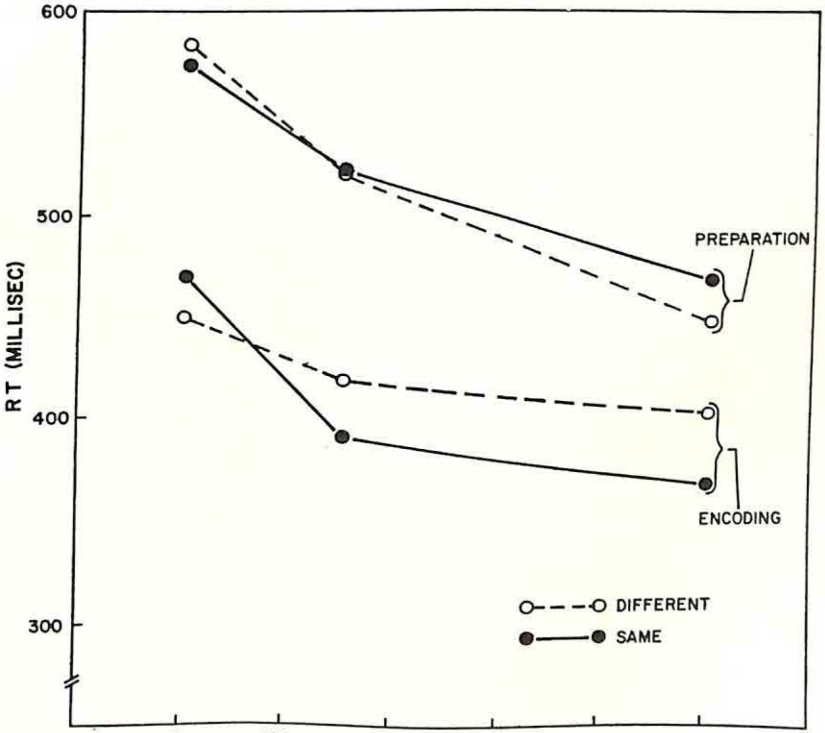


FIG. 7. "Same" versus "different" match RTs as a function of interval for preparation and encoding functions of physical match task.

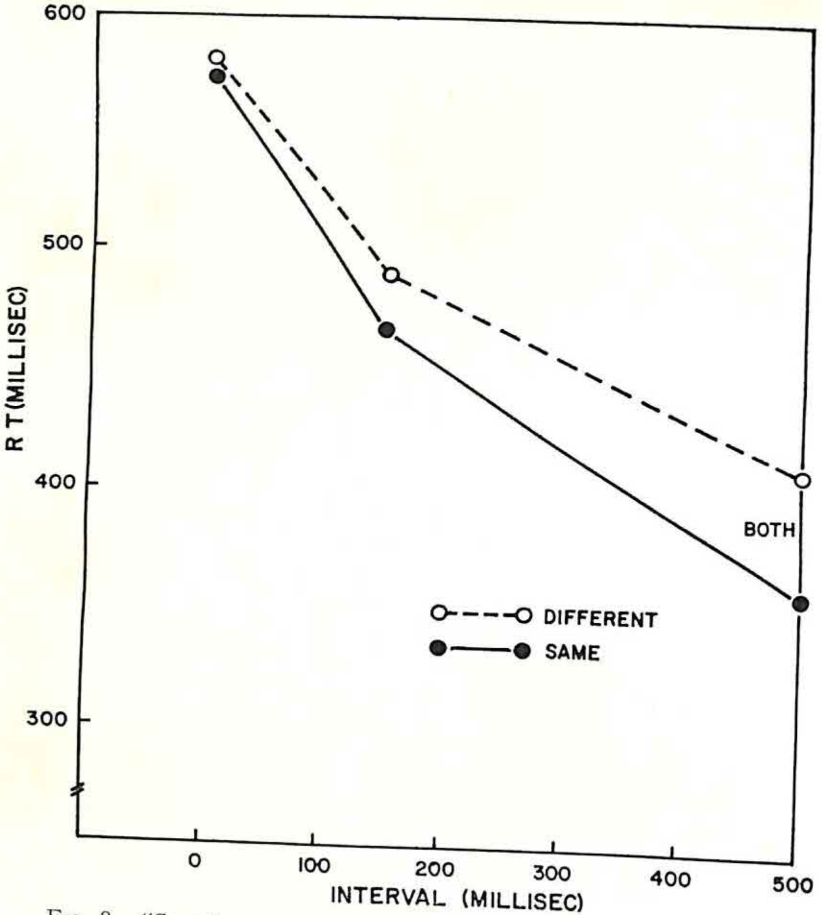


FIG. 8. "Same" versus "different" match RTs as a function of interval for "both" functions of physical match task.

a model of what *S* is looking for. When the second letter matches it, processing proceeds rapidly; if it does not, further tests are made. A similar view would be that the first letter activates internal units representing the stored information concerning that letter (Morton, 1969; Posner, 1969), which may then be more sensitive to identical stimulation for some period of time. The latter idea would presumably allow for activation of related items necessary to handle improvements at the name and vowel–consonant levels, since access to past experience is clearly needed for these matches. Both of these ideas are consistent with recent ideas about "same" and "different" responses to successive stimuli (Bamber, 1969; Tversky, 1969). What is new is that the divergence over time indicates that at least part of the advantage of "same" is due to the encoding of the first letter rather than in the criterion governing the match itself.

The results of this section suggest that encoding and preparation are separable functions with different effects on behavior. However, it appears possible to carry on both simultaneously as efficiently as either one alone. In some ways this provides an answer to Moray's question concerning whether selectivity is more efficient for an alerted organism than for an unalerted one. According to these data, whether *S* has had a previous warning signal does not change the efficiency with which he encodes a letter, at least up to the level of a name match. Nor does the relative bias in physical matching for "same" as opposed to "different" differ for alerted and unalerted conditions. These data strongly suggest that a warning signal does not improve the process of encoding information from the first letter.

So far, we have defined two separate functions having to do with attention and observed something of their relationship. It now remains to relate each of them to the third component of attention.

PROCESSING CAPACITY

One sense of attention involves competition between signals for some limited capacity system (James, 1890). The idea is that many tasks place demands on a central limited capacity system and thus tend to interfere. This basic idea has led to a number of experiments which have attempted to measure the attention demands of one task by its interference with a secondary task (Welch, 1898; Welford, 1968). One way of doing this is to use a probe as the standard secondary task. For example, Näätänen (1970) measured the evoked potential to auditory clicks, which were probes occurring during a visual RT task. Posner and Keele (1967) and Ells (1969) used a simple or choice RT task as a probe during a movement task. Foss (1969) used a probe RT to study processing of sentences. The advantage of a probe RT task is that the experimenter can measure both the RT to primary and secondary tasks and thus be certain what effects each is having on the other. In studying movements, it was found that the probe had little effect on the primary movement task, but that the probe RT reflected the central processing demands of the primary task in a very sensitive way. For example, prior to the movement, probe time was a function of a number of alternative movements and was unrelated to target width. During the movement, probe time was a function of target width and unrelated to number of alternatives (Ells, 1969). These findings suggested very strongly that probe RTs could provide a dynamic picture of the central processing demands of primary tasks.

The probe technique applied to the letter-matching sequence might indicate how preparation and encoding are related to central processing capacity. Of particular interest is whether preparation for a visual signal served to improve or retard RT to signals coming from another modality (Hernandez-Peón, 1966; Näätänen, 1970). Another question is whether encoding itself requires processing capacity in the sense of interference with a probe RT. Finally, there is the question of what aspects of letter matching seem to demand the greatest involvement of processing capacity.

Method

The same apparatus was used as described previously. A trial consisted of a warning signal followed after a WI with the first letter, which remained on until the presentation, after an ISI of a second letter adjacent to it. After each trial, feedback was provided on the correctness and RT for the letter match. The values of the WI, ISI, and intertrial interval (ITI) are given for each experiment in Table 3. On half of the trials, 50 decibels of white noise was turned on in *S*s left ear at one of eight positions in the trial. The probe occurred equally often in each of the eight positions. The *S*s were instructed that the letter task was primary and that they were to press their right index finger if the letter had the same name and their right middle finger if the names were different. They were told to press their left index finger whenever they heard the white noise burst. Feedback on the letter-matching task did not occur until both responses were complete. No feedback was provided on the noise RTs.

In Experiment I, *S*s ran six 48-trial blocks per day. Each block was identical in form. All the first letters were uppercase and the second letters

TABLE 3

INTERTRIAL, INTERSTIMULUS, AND WARNING INTERVALS AND FIRST-LETTER EXPOSURE DURATIONS (IN SECONDS) FOR EACH EXPERIMENT

Experiment	ITI	WI	Exposure duration	ISI	Probe positions (milliseconds following warning signal)							
					1	2	3	4	5	6	7	8
I	8–15 variable	1	.5	.5	−3000,	−2000,	150,	600,	1100,	1300,	1600,	1900
II	4–6 variable	.5	1	1	−500,	150,	550,	650,	800,	1000,	1650,	1800
III	4–6 variable	.5	.05 .150 .500 1.000 variable	1	−500,	150,	550,	650,	800,	1000,	1650,	1800

lowercase, so all matches were at the name level. In Experiment II, half the blocks were all uppercase (physical match) and half were identical to Experiment I (name match).

All *S*s were run individually for two days. Data from the second day only were used for those *S*s whose median probe RTs were under 1 second for all positions. Twelve *S*s were included in the analysis of each experiment (a total of three were excluded for not meeting the criterion).

Results

The basic results of each experiment are the mean of the median probe RTs at each position and the mean of the median letter match RTs at each probe position. These are shown in Figures 9, 10, 11, and 12. Differences in letter match times between physical and name matches are consistent with the instructions and suggest *S*s used a visual code for the former and the letter names for the latter (Posner, 1969).

The first point of interest is that the RT to the probe is longer when it is in the ITI than when it occurs after the warning signal. Some evidence on this point appears in Figure 9 where the two probes after the warning are slightly faster than those during the ITI. However, this improvement does not reach significance. In Experiment II, RTs at Position 2 are faster than at Position 1, but again this is not significant. However, at Positions 3 and 4, each *S* shows a definite improvement in RTs over the times obtained during the ITI ($p < .01$, by sign test). Thus, a prepared *S* gives faster RTs to the secondary task than an unprepared *S*, even though the *S* is instructed to prepare for the visual signal. An analysis of the distributions at each probe position revealed no evidence of anticipation errors (e.g., RTs to the noise of less than 200 msec.). Every effort

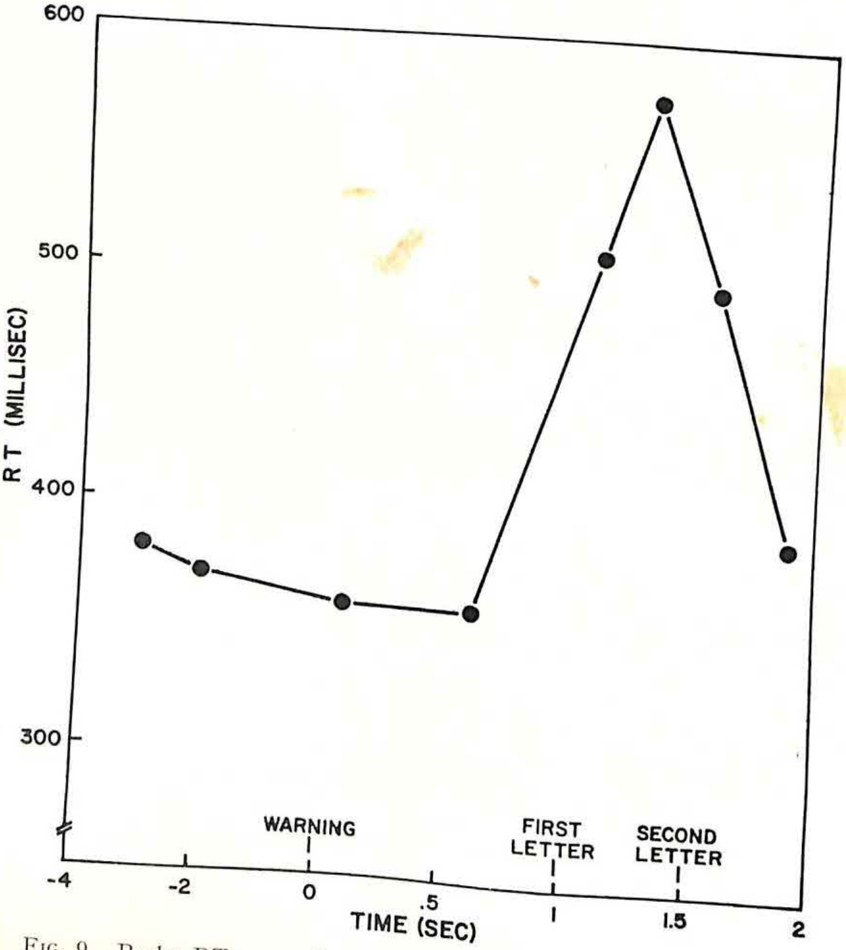


FIG. 9. Probe RTs as a function of probe position for Experiment I.

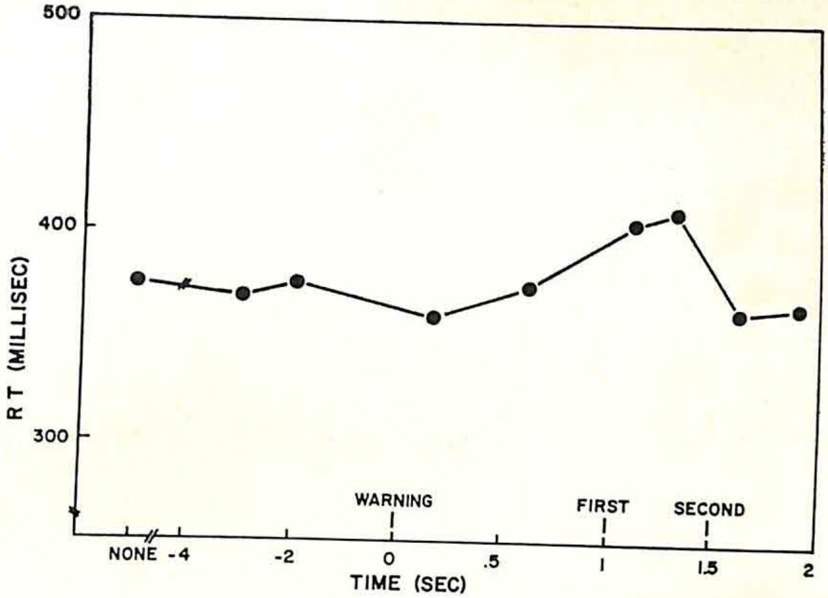


Fig. 10. Letter match RTs as a function of probe position for Experiment I.

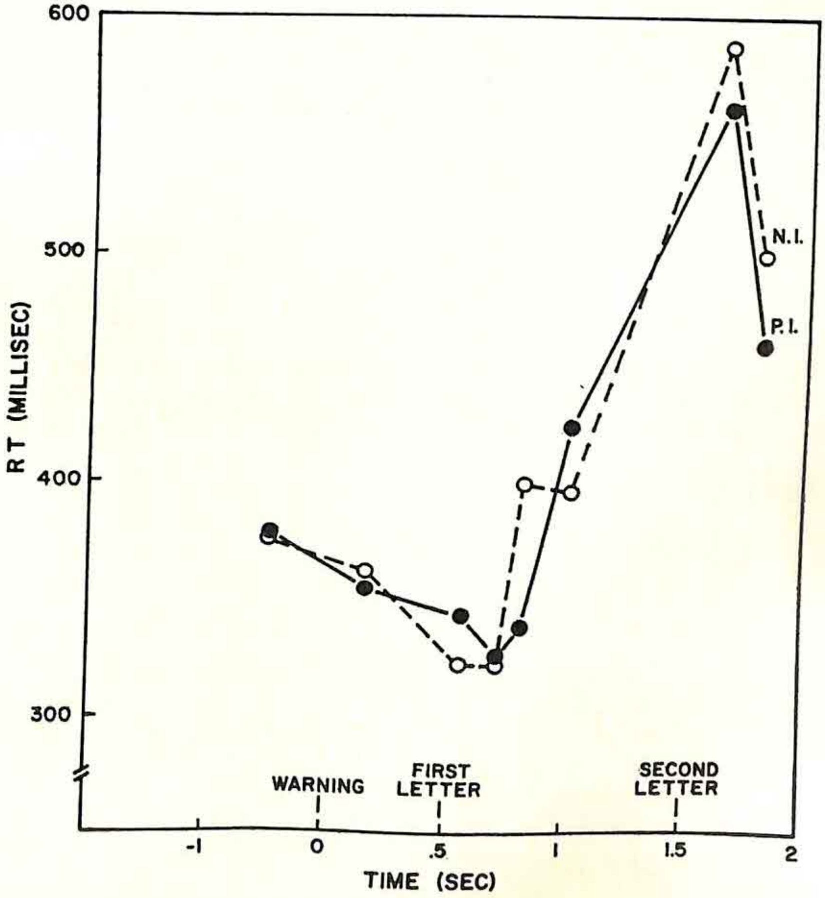


Fig. 11. Probe RTs as a function of probe position for Experiment II.

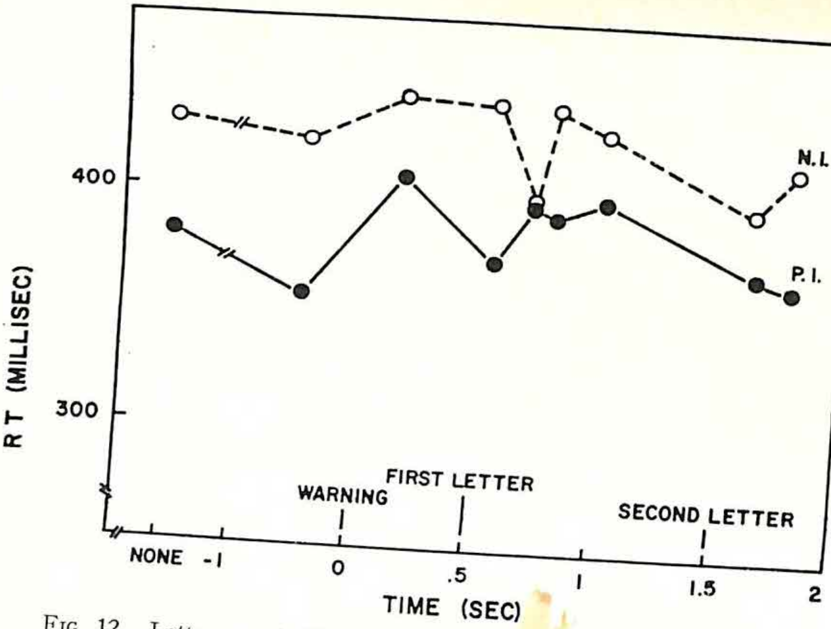


Fig. 12. Letter match RTs as a function of probe position for physical (PI) and name (NI) matches of Experiment II.

was made to insure that the preparation of *S*s was specific to the visual letter-matching task. The auditory probe occurred on only .5 of the trials and were distributed over eight different positions. All feedback involved the times for the letter task and none referred to times on the probe task.

A second point of interest is that probe times do not necessarily increase as a consequence of presenting the first letter. When the first letter was on for only .5 second (Figure 9), RT to probes after the first letter (Probe Positions 5 and 6) was increased sharply. However, when the first letter was present for 1 second (Figure 11), RT to the probe did not begin to increase until 300–500 msec. after the first letter was presented (Probe Positions 5 and 6). From previous results (see Figures 2 and 3), it would seem likely that the encoding of the first letter was pretty much complete by the time the probe began to show interference.

Third, it is clear that probe times are systematically increased at certain points. Probes which occur shortly after the second letter are always long (Figure 9, Probe Position 7, and Figure 11, Probe Positions 7 and 8). These times decrease quite consistently rom 100–400 msec. after the second letter. This result suggests that *S*s have trouble processing the probe during the response phase of the letter-match task. If probes are delayed until close to the letter-match RT, they drop back to nearly the same values as obtained during the ITI (Figure 9, Probe Position 8). Probes which occur between the two letters are increased if they fall close to the occurrence of the second letter (see Figure 11, Probe Position 6). This increase does not appear to be due to the actual appearance of the second letter. When probe distributions were computed for probes 500 msec. before the second letter in Experiment II (Figure 11, Probe Position 6), it was found that few of the RTs were long enough to occur after the second letter. Even with these excluded from the analysis, RTs to probes in this position still remained longer than those during the ITI and WI for all 12 *S*s.

Do the probe RTs really reflect the processing requirements of the letter-matching task? With the long ISI it might be possible for *S*s to switch from processing the letter to listening for the probes. In order to check on this, an experiment was run in which the first letter was turned off randomly after 50, 150, 500, or 1000 msec. The

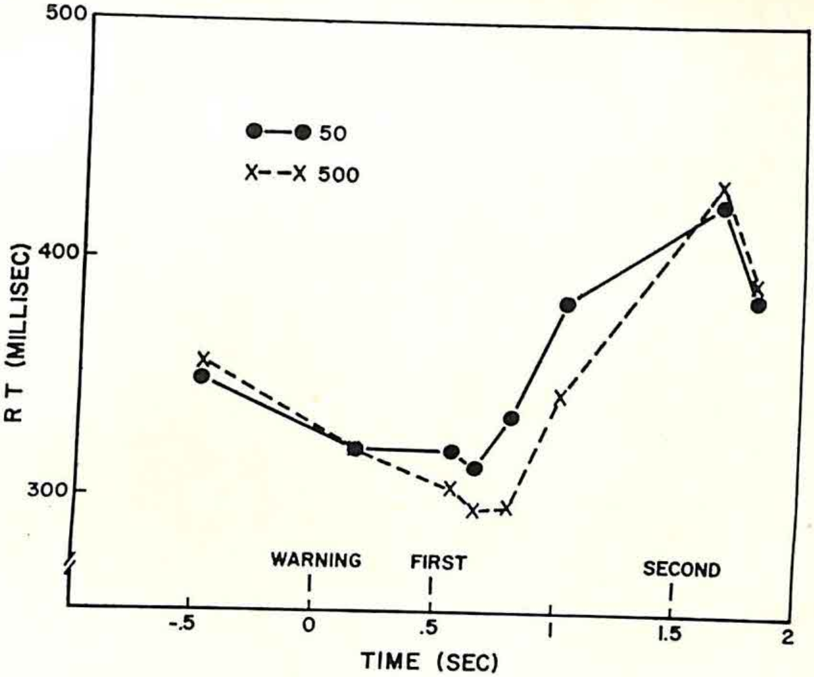


FIG. 13. Probe RTs compared for exposure durations of the first letters of 50 and 500 msec.

experiment was otherwise identical to the name letter-matching condition of Experiment II. Figure 13 shows the results for seven *S*s who ran in this experiment for five days. The data represent probe times on Days 4 and 5 for trials which had only 50-msec. exposure of the first letter and those for which the first letter was on for 500 msec. The RT results together with genally very low error rate indicates that *S*s are not switching away from the letter task. If they were, they would miss many of the 50-msec. exposures. Moreover, the probe curves from this study are quite similar to those illustrated in the previous studies. Even if the first letter was turned off after 50 msec., *S* does not show substantial interference with the probe until 500 msec.

Another possibility is that *S*s are learning the distribution of the probes and respond most rapidly to those in the middle of the distribution. This possibility seems rather remote—in part because of the low frequency of probes at any position (only on 1/16 of the trials), but also because of the complex shape of the RTs as a function of probe position. Probe Positions 4 and 5 are the two mid positions and should be fastest based on an expectancy idea. These show very different results in both Figure 9 and in the name match condition of Figure 11. Moreover, the probes at Position 8 are quite fast although they are at the end of the distribution. In addition, the results shown in Figure 13 illustrate that the point at which the first letter is turned off has small but systematic effects on the probe RTs in the direction which would be expected if *S*s were showing interference from letters which were removed somewhat more rapidly than those which remained present. Thus, it appears that probe RT reflects the attention demands of the primary task.

CONCLUSIONS

The introduction outlined three component processes involved in the study of attention. These are alertness, selectivity, and processing capacity. In this paper each of these components was identified with a particular experimental operation. Alertness was studied by varying the time between a warning signal and a pair of letters which *S* was required to match. Since the warning signal told *S* nothing about what letters he was to receive, it provided only rather minimal selective information. Selectivity was studied by providing one of the two letters so that *S* knew what he was to look for. Improvements in RT to the second letter provide a function related to encoding of the first letter. Both alertness and selectivity were found to improve performance in the letter-matching task. In both cases, the time required to reach optimal performance was about 500 msec. This value appeared to be surprisingly consistent both between the two processes and over various levels of processing (physical, name, vowel). The encoding function appeared to be somewhat steeper over the first 150 msec. than the preparation function. As the level of encoding required by the match increased, the steepness of the function appeared to be somewhat reduced.

The main consistent difference between the preparation and encoding functions was in the rate of responding to identical versus different letter pairs. The preparation interval improved physical matches and different responses equally. The encoding of one letter appeared to improve processing of physical matches more than of mismatches. This divergence between "same" and "different" responses was particularly sharp in the first 150 msec. For higher levels of instructions, no such divergence was shown.

The preparation and encoding functions are additive at least for physical and name matches. This is true even after only a 150-msec. interval when both are improving rapidly. The additivity between the two functions suggest that encoding may proceed with nearly equal efficiency regardless of the initial level of alertness. The bias introduced from the first letter appears to develop just as rapidly whether *S* has been prepared by an earlier warning signal or not. This provides one kind of answer to Moray's question about the relationship of selection and alertness. The additivity also implies that the two operations can be carried on simultaneously without interference. If, for example, *S* must both extract the name of a letter for later matching and use it as a warning signal, he may perform both functions with as great efficiency as either one alone.

This conclusion is similar to that reached by Posner and Wilkinson (1969), in which preparation, as measured by a negative shift in the electroencephalogram (contingent negative variation), was unaffected by making the warning signal a letter which had to be classified. The correspondence between the findings of the present study and previous work suggests a relationship between alertness, as used in this paper, and the contingent negative variation. If this relationship is correct, then alertness might be thought of as a general and widespread increase in cortical activation. Results of studies with the contingent negative variation support that view and suggest that this improvement is not confined merely to tasks in which there is an overt motor response (Karlin, 1970). Thus, the improvement does not appear to be confined to the motor system. How does alertness improve task efficiency? The failure of the initial level of alertness to vary the efficiency of encoding indicates that alertness does not influence the earliest stages of processing. However, it is possible that with our method, the parallel development of alertness and encoding obscures the importance of the initial state of alertness so that it does not affect the encoding function. Perhaps a more interesting possibility is that alertness varies the time when the results of the encoding process begin to occupy the limited capacity system. This would suggest the early stages of processing are unaffected by level of alertness. Studies of the relation of early components of the evoked potential to alertness (Karlin, 1970; Näätänen, 1970) are not conclusive on this point.

If preparation is viewed as a general process which is not selective, there must be other mechanisms which account for the obvious ability of *S*s to select information. The first letter appears to have a very specific effect on selection of a new letter which resembles it. If the first letter is viewed as activating a trace system related to the process of recognition (Morton, 1969; Posner, 1969), this activation may serve to make selection of already active areas of memory more efficient. Thus, in the "both" condition, a single letter would lead to a general alertness and a specific activation, which contribute additively to the overall reduction in RT. Of course, the activation of such areas of memory could come either from internal or external sources.

The nearly perfect time sharing between preparation and encoding suggests that at least one of these mental operations does not require central processing capacity. If they both did, they would be expected to interfere. In order to further test whether alertness and encoding were free of demands on processing capacity, the attention demands of the letter-matching tasks were measured by introduction of a simple probe RT task to white noise. The results suggest that a warning signal serves to reduce RT both to the main task and to the probe task, even when probes were introduced on only half the trials. Moreover, the encoding of the first letter did not seem to interfere with the probe task, provided *S*s had one second before the second letter was introduced. This was the case even when the first letter was turned off so that *S* had to encode quickly. This suggests that encoding of a letter does not require processing capacity as we have defined it here. Of course the reasonableness of the conclusion depends on the assumption that probe RT will be sensitive to processing demands of the primary task.

Interference with the probe was significant when it was introduced 500 msec. prior to the second letter. The interference was not due to the physical occurrence of the second letter, since the effect was still present when all RTs above 500 msec. were eliminated. What aspect of the letter matching causes interference with the late probes? One possible explanation is that such mental operations as rehearsal or generation of distinctive features for testing, which follow encoding of the first letter, require processing capacity, while encoding itself does not.

These results suggest that attention in the sense of central processing capacity is related to mental operations of which we are conscious, such as rehearsing or choosing a response, but is not related to the contact between the input and long-term memory that leads to the letter name. Several lines of evidence have tended to relate single-channel effects to response selection rather than all forms of stimulus processing (Herman & Kantowitz, 1970; Keele, 1970). Moreover, other research has indicated that conscious awareness is itself rather late in the sequence of mental processing (Fehrer & Raab, 1962; Libet, 1966). If single-channel effects were tied to conscious processing, one would not expect to see them closely correlated with encoding of the first letter (Deutsch & Deutsch, 1962; Norman, 1968), but rather with operations on the retrieved product of such encoding. The data obtained in this study appear to be consistent with such a view.

REFERENCES

BAMBER, D. Reaction times and error rates for "same" and "different" judgments of multidimensional stimuli. *Perception and Psychophysics,* 1969, **6**, 169–174.

BERTELSON, P. The time course of preparation. *Quarterly Journal of Experimental Psychology,* 1967, **19**, 272–279.

BOIES, S. Retention of visual information from a single letter. Unpublished master's thesis, University of Oregon, 1969.

BROADBENT, D. E. *Perception and communication.* London: Pergamon Press, 1958.

COHEN, G. Pattern recognition: Differences between matching patterns and matching descriptions to patterns. *Journal of Experimental Psychology,* 1969, **82**, 427–434.

CORBALLIS, M. C., LIEBERMAN, W., & BINDRA, D. Discriminability and central intermittency in same–different judgments. *Quarterly Journal of Experimental Psychology,* 1968, **20**, 51–61.

DEUTSCH, J. A., & DEUTSCH, D. Attention: Some theoretical considerations. *Psychological Review,* 1963, **70**, 80–90.

EGAN, J. P., GREENBERG, G. E., & SCHULMAN, A. J. Interval of time uncertainty in auditory detection. *Journal of the Acoustical Society of America,* 1961, **35**, 771–728.

EGETH, H. E. Selective attention. *Psychological Bulletin,* 1967, **67**, 41–56.

ELLS, J. Attention requirements of movement control. Unpublished doctoral dissertation, University of Oregon, 1969.

FEHRER, E., & RAAB, D. Reaction time to stimuli masked by meta contrast. *Journal of Experimental Psychology,* 1962, **63**, 143–147.

FOSS, D. J. Decision processes during sentence comprehension: Effects of lexical item difficulty and position upon decision time. *Journal of Verbal Learning and Verbal Behavior,* 1969, **8,** 457–462.

HERMAN, L. M., & KANTOWITZ, B. H. The psychological refractory period: Only half the double stimulation story. *Psychological Bulletin,* 1970, **73,** 74–88.

HERNANDEZ-PEÓN, R. Physiological mechanisms in attention. In R. W. Russell (Ed.), *Frontiers in physiological psychology.* New York: Academic Press, 1966.

JAMES, W. *Principles of psychology.* Vol. 1. New York: Holt, 1890.

KARLIN, L. Cognition, preparation, and sensory-evoked potentials. *Psychological Bulletin,* 1970, **73,** 122–136.

KEELE, S. W. The effects of input and output modes on decision time. *Journal of Experimental Psychology,* 1970, **85,** 157–164.

KOSTER, W. G. (Ed.) *Attention and performance II.* Amsterdam: North-Holland Press, 1969.

KRUEGER, L. E. Effect of bracketing lines on the speed of "same"–"different" judgment of two adjacent letters. *Journal of Experimental Psychology,* 1970, **84,** 324–330.

LA BERGE, D., VAN GELDER, P., & YELLOT, J. A cueing technique in choice reaction time. *Perception and Psychophysics,* 1970, **7,** 57–62.

LEAVITT, F. Accuracy of report and central readiness. *Journal of Experimental Psychology,* 1969, **81,** 542–546.

LEONARD, J. A. Partial advance information in a choice reaction task. *British Journal of Psychology,* 1958, **49,** 89–96.

LIBET, B. Brain stimulation and the threshold of conscious experience. In J. T. Eccles (Ed.), *Brain and conscious experience.* New York: Springer-Verlag, 1966.

MACKWORTH, J. A. *Vigilance and habituation.* London: Penguin Press, 1970.

MORAY, N. *Listening and attention.* London: Penguin Press, 1970.

MORTON, J. Interaction of information in word recognition. *Psychological Review,* 1969, **76,** 165–178.

NÄÄTÄNEN, R. Evoked potential, EEG, and slow potential correlates of selective attention. In A. F. Sanders (Ed.), *Attention and performance III.* Amsterdam: North-Holland Press, 1970.

NICKERSON, R. S. Response times for same-different judgments. *Perceptual and Motor Skills,* 1965, **20,** 15–18.

NORMAN, D. A. Toward a theory of memory and attention. *Psychological Review,* 1968, **75,** 522–536.

POSNER, M. I. Abstraction and the process of recognition. In G. Bower & J. T. Spence (Eds.), *Psychology of learning and motivation.* Vol. 3. New York: Academic Press, 1969.

POSNER, M. I., & KEELE, S. W. Attention demands of movement. In, *Proceedings of the 16th International Congress of Applied Psychology.* Amsterdam: Swets & Zeitlinger, 1967.

POSNER, M. I., BOIES, S. J., EICHELMAN, W. H., & TAYLOR, R. L. Retention of visual and name codes of single letters. *Journal of Experimental Psychology,* 1969, **79** (1, Pt. 2).

POSNER, M. I., & MITCHELL, R. F. Chronometric analysis of classification. *Psychological Review,* 1967, **74,** 392–409.

POSNER, M. I., & WILKINSON, R. T. On the process of preparation. Paper presented at the meeting of the Psychonomics Society, November 1969.

REYNOLDS, D. Effects of double stimulation: Temporal inhibition of response. *Psychological Bulletin,* 1964, **62,** 333–347.

SANDERS, A. F. (Ed.) *Attention and performance I.* Amsterdam: North-Holland Press, 1967.

SANDERS, A. F. (Ed.) *Attention and performance III.* Amsterdam: North-Holland Press, 1970.

SMITH, E. E., & NIELSEN, G. D. Representations and retrieval processes in short-term memory: Recognition and recall of faces. *Journal of Experimental Psychology,* 1970, **85,** 397–405.

SOKOLOV, Y. N. *Perception and the conditioned reflex.* London: Pergamon Press, 1963.

STERNBERG, S. The discovery of processing stages: Extension of Donders method. In W. G. Koster (Ed.), *Attention and performance II.* Amsterdam: North-Holland Press, 1969.

TRIESMAN, A. Strategies and models of selective attention. *Psychological Review,* 1969, **76,** 282–299.

TVERSKY, B. Pictorial and verbal encoding in a short-term memory task. *Perception and Psychophysics,* 1969, **6,** 225–233.

WELCH, J. On the measurement of mental activity through muscular activity and the determination of a constant of attention. *American Journal of Physiology,* 1898, **1,** 288–306.

WELFORD, A. T. *Fundamentals of skill.* London: Methuen, 1968.

WILKINSON, R. T., & HAINES, E. Evoked response correlates of expectancy during vigilance. *Acta Psychologica,* 1970, **33,** 402–413.

(Received November 4, 1970)

Psychological Review
1971, Vol. 78, No. 5, 409–419

STATISTICAL DECISION MODEL FOR AUDITORY WORD RECOGNITION[1]

JOHN R. FREDERIKSEN[2]

Brandeis University

A statistical decision model for auditory word recognition is proposed which accounts for response dependencies in a multitrial recognition task. An experimental test of the model is described in which the effect of increasing stimulus word frequency is to increase the criterion level of mismatch used in rejecting hypotheses. Additional data are presented which suggest that (*a*) hypothesis word frequency is correlated with stimulus word frequency, and (*b*) high-frequency words may be phonetically distinct from low-frequency words. The implications of these findings for "sophisticated guessing" and decision theoretic explanations of the word-frequency effect are discussed.

It is a well-known finding that words which are common in the language are more readily identified than are words which are uncommon in their occurrence (Howes, 1954; Howes & Solomon, 1951). While this word frequency effect has become a relatively routine observation in studies of auditory and visual word recognition, an adequately documented explanation for the phenomenon has as yet eluded investigators in the area. Until quite recently, the most successful accounts have explained the word frequency effect as the result of a strategy which the subject (*S*) uses in guessing the correct word on those trials where the stimulus word is not correctly perceived (e. g., Pollack, Rubenstein, & Decker, 1960; Savin, 1963; Solomon & Postman, 1952). In this "sophisticated guessing" model, it is assumed that even when *S* has incorrectly perceived the stimulus word, he has still received enough information about phonetic characteristics of the word to rule out all but a restricted set of alternatives from which he chooses a word at random to serve as his guess. In addition, it is assumed that the selection of a response from the pool of available alternatives is carried out with a bias in favor of choosing words of high frequency ahead of words of low frequency. According to this theory, the word frequency effect arises from the greater likelihood of *S*'s correctly guessing a high-frequency stimulus on those occasions when he has not correctly perceived it.

This view of the word recognition process has been challenged by Broadbent (1967), who proposes instead that the word frequency effect is due to a criterion bias, as conceived in signal detection theory. According to the signal detection formulation, the presentation of a stimulus word evokes, to some extent, every word in *S*'s repertoire as a potential response. Associated with each possible response word is a decision axis for the particular response alternative in question. Information gained from the presentation of a stimulus (the "observation") determines a point on the decision axis associated with each word in *S*'s repertoire. The "observation" corresponds to phonetic and structural information about the word which has been perceived, and a high value on the decision axis corresponds, therefore, to a condition of phonetic-acoustic match between potential response and perceptual input, while a low value on the decision axis corresponds to an acoustic mismatch between input and response word. According to Broadbent's theory, *S* must decide among the alternative words in his repertoire by comparing their values on the decision axis for the available perceptual input and by selecting the word with the highest value, that is, the one which provides the best match to the stimulus. This decision is complicated, however, by the fact that *S* is biased in the direction of accepting lower values on the decision axis for high-frequency words than for low-frequency words as a criterion for responding. This assumption is justifiable in light of the fact that high-frequency words have higher a priori subjective probabilities of occurring in the context of the experiment than do low-frequency words. The resulting criterion bias leads *S* to favor high-frequency hypotheses over low-frequency hypotheses when their values on the decision axis are the same. The *S* will thus respond more readily with the correct word when it is a word of high frequency than when it is of low frequency. Broadbent's attempt to invalidate the sophisticated guessing theory by looking at data on the word frequencies of errors made in response to high- and low-frequency stimuli has been disputed in a recent paper by Catlin (1969),[3] who argues that a modification in the interpretation of the effects of acoustical equivalence of high- and low-frequency words yields a guessing model which can account for Broadbent's data. In addition, Morton (1968) has shown that both the sophisticated guessing and the decision theory models require revision in order to account for the data presented by Brown and Rubenstein (1961), so that at present no clear empirical basis exists to distinguish between the two points of view. In the present study, an attempt was made to demonstrate criterion bias in auditory word recognition, using mathematical and experimental techniques which allow us to provide a metric for the decision axis and to actually locate the distributions for response words and the differential criteria which, it is hypothesized, are used for high- and low-frequency words.

---

[1] This work was begun while the author was a Research Associate at the Department of Psychology, Massachusetts Institute of Technology. Support was received from the National Institutes of Health Training Grant 2T01 GM-01064-06 to H.-L. Teuber, from National Science Foundation Institutional Grant GU 2839, and from National Institutes of Health Institutional Biomedical Grant No. 7044-04 to Brandeis University. Material in this paper was originally presented at the meeting of the Eastern Psychological Association, Atlantic City, April 1970.

[2] Requests for reprints should be sent to John R. Frederiksen, Department of Psychology, Brandeis University, Waltham, Massachusetts 02154.

A DECISION MODEL FOR SEQUENTIAL WORD RECOGNITION

In order to investigate the effect of word frequency on factors determining *S*'s choice of response in auditory word recognition, a sequential recognition procedure was employed. In such a task, *S* is given an opportunity to make a series of consecutive guesses about the identity of a masked stimulus word, as the degree of masking is reduced (and the signal-to-noise ratio [S/N] increased) over a series of trials (Frederiksen, 1967, 1969). For example, a stimulus word might be presented 15 times, starting at an S/N of −6 decibels (db.) on the first trial, and with an increase in S/N of 1 db. on each subsequent trial. After each trial, *S* writes down his guess as to the identity of the masked word.

In the sequential task, *S*'s response on Trial $i+1$ is clearly dependent not only on the acoustical character of the stimulus on that trial, but on his response on the previous trial (i) as well. Since *S* is required to write down his current hypothesis after each trial, response statistics, such as the trial on which he first writes the correct word, can be computed. Through an analysis of response statistics, obtained over a series of recognition problems, estimates can be obtained of parameters which reflect the underlying decision processes operating in perceptual recognition.

---

[3] Catlin's argument follows the urn model for sophisticated guessing of Pollack et al. (1960), which assumes that the effect of a stimulus presentation is to select a sample of balls from an urn representing *S*'s response hierarchy. All of the balls in the sample selected correspond to words which bear a phonetic relationship to the stimulus word, as it has been in part perceived. While a portion of the balls in this "restricted" urn correspond to the stimulus word, the remaining balls represent erroneous responses. Catlin maintains that the assumption of an acoustical equivalence of high- and low-frequency words implies that the number of balls in the restricted urn corresponding to high-frequency (E_H) and low-frequency (E_L) errors does *not* depend on the frequency of the stimulus word. From this assumption, he arrives at the conclusion that the proportion of high-frequency errors $E_H/(E_H+E_L)$ is independent of stimulus word frequency, a result which is contrary to Broadbent's conclusion. It might here be pointed out that Catlin's assumption that E_H and E_L are independent of stimulus word frequency implies also that the size of the sample of balls drawn from the original urn is a function of stimulus word frequency, since the restricted urn always contains balls representing the stimulus word, and the number of balls corresponding to the stimulus word depends on the stimulus word frequency. Whether this implication of Catlin's assumption is consistent with the notion that the presentation of a stimulus effects a random selection of balls from the initial urn is a question worth pondering.

The Decision Model

In the terms of statistical decision theory, we suggest that *S*'s response on any particular trial is determined by his choice of a criterion used in rejecting his previous response. On each trial, *S* is thought to compare his current hypothesis with information available from the masked stimulus, and to test for a disparity between the two. Since the disparity between hypothesis and perceptual input to the decision system varies from moment to moment, it projects a distribution of values on a judgmental (match–mismatch) axis, as in the top drawing in Figure 1. When the degree of mismatch on a particular trial *exceeds* a criterion level, the *S* rejects his present response and reenters his response hierarchy to select, on perceptual grounds, a new response. If, on the other hand, the degree of mismatch fails to reach the criterion level, the *S* instead perseverates with his previous response. The area under the probability density function to the left of the criterion corresponds to the probability (α) that *S* will perseverate in this way.

The second drawing in Figure 1 shows one way in which two hypotheses might differ, namely, in their positions on the decision axis. Two distributions are shown, one for a tenable hypothesis (Hypothesis 1) and the other for one that is less so (Hypothesis 2). Since the average degree of mismatch is higher for Hypothesis 2 than for Hypothesis 1, the probability α is smaller for the second hypothesis than for the first, indicating that an *S* currently holding the second hypothesis will be more likely to reenter his response hierarchy and select a new response than will an *S* holding the first one. One experimental variable which might alter the expected degree of hypothesis mismatch on successive trials is the step-size change in S/N occurring on each trial. A large reduction in noise level from one trial to the next would presumably reveal more information about the discrepancy between stimulus and hypothesis than would a small change in S/N.

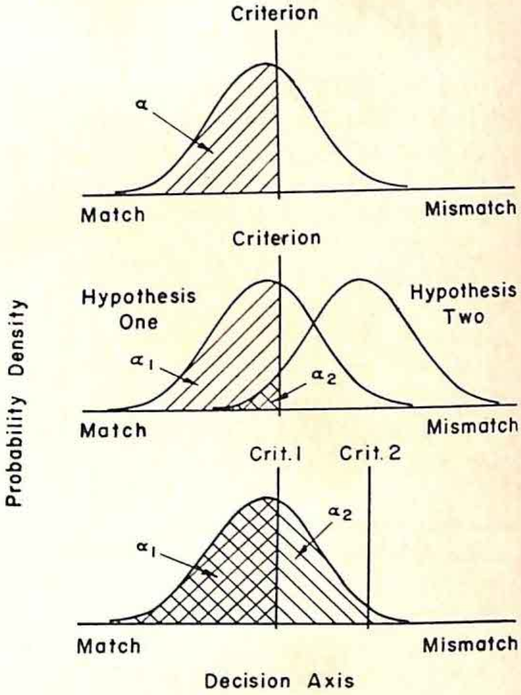


FIG. 1. An illustration of the effects of stimulus-hypothesis congruence and criterion shift on α, the probability of perseveration, as it is represented in the decision model.

The drawing at the bottom of Figure 1 illusrates a second way in which two hypotheses may differ, namely with regard to the criteria which are used by *S* in his decision to reject them. Under the circumstances, even though the hypotheses are equal in their degree of match to the perceptual input, *S* will reject one more readily than the other. In particular, following Broadbent's (1967) suggestion, Criterion 1 in the figure might correspond to the criterion for rejecting a low-frequency hypothesis word, while Criterion 2 might correspond to that used in rejecting a high-frequency word. The *S* would then be biased toward accepting less discrepant information as grounds for rejecting the low-frequency hypothesis than he would in order to reject a hypothesis of high frequency.

The key parameter in each of these illustrations is α, the probability of perseverating with one's previous hypothesis on a given recognition trial. Direct estimation of the value of this parameter for real data (by simply counting perseverative responses) is complicated by the fact that an S can emit the same response on two consecutive trials for either of two reasons: (*a*) he can refuse to reject his previous hypothesis when its degree of mismatch has failed to exceed his criterion, or (*b*) he can resample his previous response after having first rejected it, since sampling from the response hierarchy is carried out with replacement (cf. Frederiksen, 1969). While direct measurement of α is thus ruled out, indirect measurement of α is possible through the application of a mathematical model for the word recognition process.

The Mathematical Model

In the mathematical model, it is assumed that whenever S rejects a previous hypothesis, he reenters his response hierarchy to select, with replacement, a new response. His probability of selecting the correct response is assumed to be the value one would predict in the absence of intertrial response dependencies; it thus will depend on S/N but *not* on the number of previous trials to which S has been exposed. The probability that the new response is *correct* is assumed, therefore, to be a linear function $\gamma + \theta s_i$ of the S/N s_i on the particular trial (i). The parameter γ represents the probability that S responds correctly at the reference S/N which is the S/N on Trial 1. Theta represents the change in probability correct for a unit change in S/N. Ample evidence supporting this linearity assumption for situations in which successive trials are independent exists in the literature and here needs no detailed citation (e.g., Hawkins & Stevens, 1950; Miller, 1947). There certainly is no reason to suspect great departures from linearity for the range of S/N's employed in the present study (–6 db. to +8 db.). In any event, if serious departures from the model predictions are encountered, alternative functions relating probability correct to S/N could be considered, such as the logistic or the normal ogive. The function $\gamma + \theta s_i$ can be thought of as giving the probability of S's being correct on Trial i that would be expected if S's response on this trial were independent of his response on Trial $i - 1$.

The trial-to-trial response dependencies, which are more characteristic of a sequential recognition task, are dealt with in the model by assuming that there is a fixed probability α (for a given set of experimental conditions) that S perseverates with his previous response on any given trial, rather than reenter his response hierarchy to select a new response (with probability $1 - \alpha$). This assumption, combined with the assumption of linearity, is sufficient to derive a theoretical probability distribution of first recognition trials obtained in a group of Ss, for a set of equivalent items. A first recognition trial is defined as the trial on which S states for the first time the correct response word. Letting P_j stand for the probability that S is correct for the first time on Trial j, and setting $s_j = 0$ for $j = 1$ (which merely establishes the S/N on Trial 1 as the reference level), we can write

$$P_1 = \gamma \qquad [1]$$

for S's probability of being correct for the first time on Trial 1;

$$P_2 = (1 - \gamma)(1 - \alpha)(\theta s_2 + \gamma) \qquad [2]$$

for his probability of being correct for the first time on Trial 2; and

$$P_j = (1 - \gamma)(1 - \alpha)(\theta s_j + \gamma) \times \prod_{i=2}^{j-1} [1 - (1 - \alpha)(\theta s_i + \gamma)] \qquad [3]$$

for his probability of being correct on Trial j for the first time. Finally, the probability P_f that S fails to recognize at all during the k trials of the experiment is given by

$$P_f = (1 - \gamma) \times \prod_{i=2}^{k} [1 - (1 - \alpha)(\theta s_i + \gamma)]. \qquad [4]$$

Together, Equations 1 through 4 specify the theoretical probability distribution of first recognition points in the sequential recognition task. It should be noted that since we have chosen to fit our mathematical model to the distribution of first recognition trials, no data beyond the trial on which *S* first writes the correct response word are considered in the analysis. Admittedly, information is lost in using this response statistic, since *S*'s may reject the correct response on the next trial and then return to it later. However, the advantage in using this response statistic lies in the relative simplicity of the equations for the probability distribution and in the experimental independence of α from the word frequency of the stimulus. The details of the derivation of the above equations and of the maximum likelihood estimation procedure are presented elsewhere (Frederiksen, in preparation). Fitting the theoretical probability distribution to experimental data provides, in addition to estimates of γ and θ, an estimate of the probability α that *S* perseverates with his current response on any given trial. Maximum likelihood estimates of the three model parameters can be routinely calculated for any such distribution, using a computer program that employs an iterative solution for maximizing the log likelihood function.[4] Note again that it is through the inclusion of a perseveration parameter α that the mathematical theory has taken into account the dependency of behavior on Trial $i+1$ on that of Trial *i*. The mathematical model, however, makes no assumptions about the sensory–perceptual basis of the decision to perseverate. It is only in the terms of the decision model that the probability α is interpreted as the likelihood that the degree of mismatch between hypothesis and acoustic (perceptual) input fails to exceed a criterion level. The mathematical model thus has here the status of a measurement model, which in its application enables us to study probability of perseveration as it relates to other characteristics of the experimental situation.

AN EXPERIMENTAL TEST OF THE DECISION MODEL

The present experiment was designed to test some implications of the decision model for sequential word recognition outlined above. Experimental manipulations were sought which would vary (*a*) the position of the hypothesis distribution on the judgmental (match–mismatch) axis, and (*b*) the criterion employed by *S* in rejecting his previous hypothesis. In order to satisfy Requirement *a*, stimuli were presented using each of three different rates of change in S/N: 0 db. per trial, 1 db. per trial, and 2 db. per trial. It was hypothesized that the distributions obtained for these three conditions would be similarly ordered in their positions on the decision axis.

In order to influence *S*'s choice of criterion as specified in Requirement *b*, stimulus words belonging to two word frequency classes were presented, in the belief that manipulating stimulus word frequency would have the effect of altering hypothesis word frequency in a similar direction. Such a correlation between stimulus and hypothesis word frequency is not without experimental support (cf. Broadbent, 1967; Brown & Rubenstein, 1961), and data necessary to test for this effect would certainly be available in the results of the present study. Accordingly, anticipating that this correlation between stimulus word frequency and hypothesis word frequency would be found to occur, it was further hypothesized that the criterion used by *S* in rejecting hypotheses formed in response to high-frequency words would exceed the criterion used in rejecting hypotheses formed in response to low-frequency words. Finally, note that while this second hypothesis resembles the assumption underlying Broadbent's explanation for the word frequency effect, we are here dealing with criteria used in *rejecting* previously held hypotheses, whereas Broadbent asserts that differential criteria are employed in the *selection* of new responses. However, since the rationale underlying the present hypothesis involves the notion that high-frequency words have a higher subjective probability of occurring in the context of the laboratory experiment than do low-frequency words, and since this notion appears to be as applicable to response selection as it is to response rejection, a verification of the second hypothesis should have some bearing on our

---

[4] Copies of the FORTRAN program, written for the IBM 360, are available from the author.

TABLE 1

RESULTS OF THE EXPERIMENT DEALING WITH CRITERION SHIFT IN AUDITORY WORD RECOGNITION

Item	$s_{j+1} - s_j$					
	0 db. per trial		1 db. per trial		2 db. per trial	
	Word frequency					
	High	Low	High	Low	High	Low
Parameter						
Alpha	.730	.509	.533	.250	.413	.000
Gamma	.050	.034	.141	.033	.077	.024
Theta	.000	.000	.008	.012	.036	.015
Log L	−15.471	−17.420	−30.448	−34.211	−20.143	−18.804
χ^2	—	—	6.167	16.272	6.631	5.607
df	0	0	9	10	5	4
p	—	—	$<.75$	$<.10$	$<.25$	$<.25$
z	.62	.03	.08	−.68	−.22	−1.64[a]

[a] Since an inverse normal transformation for $p = 0$ is indeterminate, alpha was arbitrarily set to .05 for the purpose of computing a *z* score.

judgment about the plausibility of Broadbent's model for the selection of responses.

Method

Subjects. The *S*'s in this experiment were 10 undergraduates attending Brandeis University. Each *S* volunteered his services and was paid $2.00 for his participation in the study. Those *S*s who reported a history of hearing defects were excluded from the sample.

Procedure. In the present experiment, the *S*s were presented a set of 72 masked disyllabic stimulus words to recognize. Each word was presented for a series of consecutive trials, and after each recognition trial, the *S*'s task was to write down his guess as to the identity of the masked word. The 2 × 3 factorial design called for two levels of Thorndike-Lorge frequency (high and low) and three rates of change in S/N (0 db. per trial, 1 db. per trial, and 2 db. per trial). The S/N on Trial 1 was in all cases — 6 db. The total number of trials for the three rates of change in S/N were 10, 15, and 8, respectively. The order of presentation of the six conditions was randomized so that within any block of 12 recognition problems, two instances of each condition occurred. The 12 words assigned to each combination of word frequency and rate of change in S/N were randomly selected from the Thorndike-Lorge tables (Thorndike & Lorge, 1944) subject to the following constraints: (*a*) Words having Thorndike-Lorge frequencies (general count) above 27 were classified as high frequency, with all other words classified as low frequency; (*b*) all words were of two syllables; (*c*) words assigned to the six conditions were balanced as to location of the stressed syllable; (*d*) each word was chosen to have at least one phonetically similar word of the same frequency class available. Stimuli were spoken by a student radio announcer and were recorded on magnetic tape using an Ampex Model PR-10 tape recorder. Reproduction was by means of a Magnacorde Model 1024 tape recorder and a Telex St-11 Stereo Headset.

Each *S* was presented with the same series of 72 word recognition problems along with 12 practice problems. The loudness of the stimulus was adjusted by each *S* to a comfortable listening level before the experimental series was begun. Then, *S* pressed a key to begin each recognition problem, and through suitable control circuitry, the Magnacorde tape deck was programmed to halt at the conclusion of each problem. The *S*s thus paced themselves and were allowed to rest whenever they desired to do so. Trials within a problem, however, were presented at a fixed rate of one trial every four seconds.

Results

By computing maximum likelihood estimates of the model parameters for each of the six conditions, separate estimates of α, γ, and θ were obtained for each of the experimental conditions. These results are presented in Table 1. The fit of the mathematical model can be evaluated for the 1 and 2 db. per trial data using a chi-square test of goodness of fit, and can be seen to be satisfactory in each case (with p taking on values between .75 and .10). Comparing the γ's obtained for high- and low-frequency words for each rate of change in S/N, it can be seen that in every case, the high-frequency words are recognized more frequently at the reference S/N (− 6 db.) than are the low-frequency words. Thus, the word frequency effect is evident in these data. At the same time, there is no consistent effect of word frequency on the slope parameter θ, which suggests that there is no perceptual advantage of one kind of word over another.

The present model enables us to directly test the criterion-shift theory of the word frequency effect. It is clear that if hypothesis word frequency is correlated with stimulus word frequency, lowering the average frequency of the stimulus words should result in a decrease in the probability α that the level of mismatch fails to exceed criterion, due to a criterion shift in the direction of increasing the readiness of the *S* to reject hypotheses. Increasing the average word frequency, however, will have the opposite effect—that of increasing the probability α that *S* will perseverate with his former response. On the other hand, increasing the step-size change in S/N occurring on each trial (at a fixed level of word frequency) should result in a decrease in α, since the average degree of mismatch should be greater when the change in noise level is considerable than when it is slight. The experimental results clearly support both of these predictions. For each rate of change in S/N, the estimate of α for high-frequency words is greater than it is for low-frequency words. Moreover, the effect of increasing the step-size change in S/N was to increase the average discrepancy between hypothesis and perceptual input to the decision process, with a consequent decrease in the value of the parameter α.

Since the α parameters obtained for high- and low-frequency words (for a fixed rate of change in S/N) can be interpreted as cutting off two areas of a normal distribution, it is possible to solve for the mean and standard deviation of this distribution (and in the process establish a scale for the decision axis) if we set the origin of the scale at the low-frequency criterion and define the unit of measurement as the difference between the high- and low-frequency criteria. If we assume in addition that the difference between the criteria for high- and low-frequency words is independent of the step-size change in S/N, a metric has then been specified for the decision axis because we then have a common unit. The three distributions corresponding to the 0, 1, and 2 db. per trial

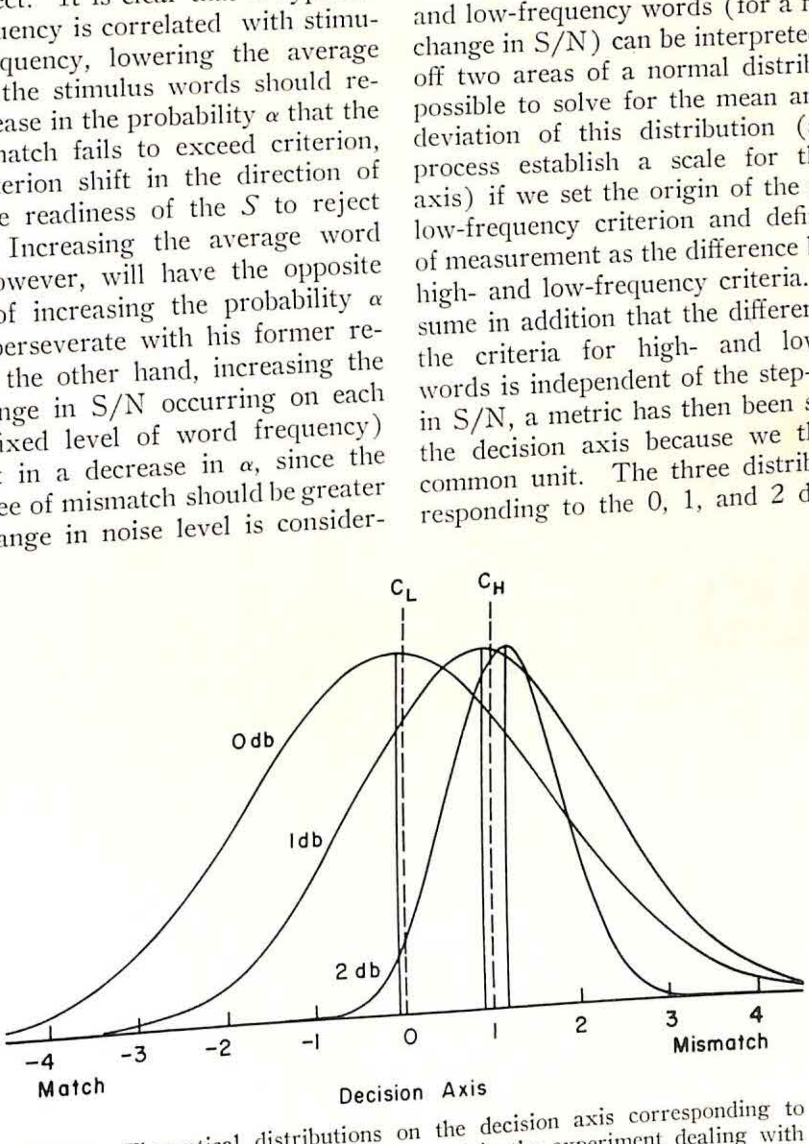


FIG. 2. Theoretical distributions on the decision axis corresponding to the three rates of change in S/N employed in the experiment dealing with criterion shift in word recognition. (The two criteria are for high-frequency words [C_H] and low-frequency words [C_L]. The origin was arbitrarily set at C_L, and the unit was chosen to be C_H–C_L.)

conditions can then be located on this common scale. If z_H denotes the z score corresponding to the probability α obtained for the high-frequency words and if z_L is the z score corresponding to α obtained for the low-frequency words, then the mean of the distribution is given by

$$\mu = \frac{-z_L}{z_H - z_L},$$

while the standard deviation of the distribution is given by

$$\sigma = \frac{1}{z_H - z_L}.$$

The relative spacing and widths of these distributions for 0, 1, and 2 db. per trial conditions are shown in Figure 2. The ordering of the three distributions is in the hypothesized configuration and is taken as support for the interpretation that (*a*) the expected degree of mismatch is positively related to the amount of change in noise level, and (*b*) differential criteria for rejection of hypotheses are employed for high- and low-frequency words.

DISCUSSION

The results of this experiment are clearly consistent with predictions derived from the decision theory model, and lend support to the notion that differential criteria are associated with words belonging to different frequency classes. The present findings show in particular how such multiple criteria are used in making perceptual decisions in the context of a sequential word recognition task. They show how the amount of discrepant information necessary for rejecting high-frequency responses exceeds that necessary for rejecting low-frequency hypotheses. This difference is comparable in magnitude to that produced by increasing the step-size change in S/N from 0 db. per trial to 1 db. per trial (see Figure 2). These results are also consistent with Broadbent's criterion-placement model for response selection, although they do not necessarily imply that such a model is correct. Unfortunately, the acceptance of a single model for response selection must await further research. However, as will be seen from the data on word frequencies of errors, it is possible that phonetic correlates of word frequency may play a part in leading *S* to select responses which belong to the same frequency class as that of the stimulus. This finding, in turn, has implications for a choice of response-selection model.

TABLE 2

PROPORTIONS OF HYPOTHESIS WORDS IN EACH OF THREE WORD FREQUENCY RANGES

Stimulus word frequency	Hypothesis word frequency			Total
	AA (>100)	100–21	20–1	
AA (>100)	.336 (66)	.328 (64)	.336 (66)	(196)
100–21	.266 (154)	.334 (194)	.400 (232)	(580)
20–1	.216 (144)	.310 (206)	.474 (315)	(665)
Total	.253 (364)	.322 (464)	.425 (613)	(1441)

Note.—Raw frequencies are given in parentheses.

In order to understand more clearly why *S*s set a higher criterion for rejecting hypotheses developed in response to high-frequency stimuli than to low-frequency stimuli, a tally was made (and is shown in Table 2) of the number and proportion of hypotheses falling in each of three Thorndike-Lorge frequency ranges for stimulus words in each of the same three frequency ranges. Note that in tallying the hypothesis word frequencies, correct responses were excluded from consideration. It is evident that the majority of hypotheses developed in response to low-frequency stimuli belong in the low-frequency class, while equal numbers of high-, medium-, and low-frequency hypotheses are developed in response to high-frequency stimuli. This result is clearly inconsistent with Broadbent's (1967) conclusion that errors of high frequency are in a constant ratio (regardless of stimulus word frequency) to errors of low frequency. It should be pointed out again that the association illustrated in Table 2 between stimulus word frequency and hypothesis word frequency has been found previously in at least two studies. Brown and Rubenstein (1961) reported a positive relationship between the frequency class of the stimulus and that of responses to the stimulus; Broadbent (1967) observed a trend similar to this in his disyllabic data, but not in his data for monosyllabic words.

The relation observed here between the word frequency of stimulus and that of hypothesis is also at variance with predictions from sophisticated guessing theory, as Broadbent (1967) has pointed out. Evidently, while the effect of the stimulus presentation is to delineate a restricted set of alternatives from which *S* chooses his response, this restricted set of potential responses is far from being random with respect to word frequency. The above data indicate that the set contains a disproportionate number of words belonging to the same frequency class as the stimulus. An implication of this finding, as Brown and Rubenstein (1961) pointed out, is that stimulus words carry information as to their frequency class even when presented with an accompanying masking sound.

The correlation obtained between stimulus word frequency and hypothesis word frequency suggests, then, the additional theoretical hypothesis that low-frequency words might be characterized by particular phonetic features which are common only to infrequently occurring words. As a preliminary test of this "phonetic hypothesis," a linguistic comparison was made of the high- and low-frequency stimulus words employed in the present experiment. In particular, counts were made of the number of occurrences of Zipf's (1935) "high-complexity" and "low-complexity" phonemes. These two groups of phonemes, Zipf argued, differ in articulatory difficulty, and on the basis of "the principle of least effort," should be (and are) similarly ordered with respect to overall frequency of occurrence in the spoken language. Zipf's figures for the percentage of frequency of occurrence of each of eight phonemes in spoken English are reproduced in the first column of Table 3. The eight phonemes include the unvoiced and voiced stops, and the two nasals n and m. Also given in Table 3 is a new set of within-word-class frequency counts, in which the number of occurrences of each phoneme have been tallied within each of the two classes of words, belonging to different frequency groups. In every case, there is an *increase* in the percent frequency of occurrence of high-complexity phonemes (d, b, g, and m) as one goes from high-frequency words to low-frequency words. No such consistent increases are found for low-complexity phonemes (t, p, k, and n)—a result which is exactly as Zipf would have predicted.

TABLE 3

FREQUENCIES OF OCCURRENCE OF ZIPF'S HIGH- A
LOW-COMPLEXITY PHONEMES IN STIMULUS
WORDS OF TWO FREQUENCY LEVELS

Phoneme type	% frequency of occurrence in English speech[a]	High-frequency stimulus words		Low-frequency stimulus words	
		Frequency	%	Frequency	%
Total phonemes	—	181	—	192	—
t	7.13	11	6.1	10	5.2
p	2.04	6	3.3	4	2.1
k	2.71	1	.6	4	2.1
Total unvoiced	11.88	18	9.9	18	9.4
d	4.31	5	2.8	10	5.2
b	1.81	1	.6	4	2.1
g	.74	3	1.6	6	3.1
Total voiced	6.86	9	5.0	20	10.4
n	7.24	8	4.4	9	4.7
m	2.78	4	2.2	7	3.6

[a] Data are taken from Zipf (1935, pp. 75 and 79).

The usual statistical test of the equality of proportions cannot be applied to the data in Table 3, since it is unreasonable to assume that there is an equal chance (under the null hypothesis) for each phoneme within a word to belong to a particular phoneme class. In order to obtain an indication of the statistical reliability of the difference in frequencies of occurrence of high- and low-complexity phonemes for the two classes of words, it was decided instead to look at the proportions of words of high and low frequency that contain at least one phoneme from a given set. These results are given in Table 4.

If we compare the proportions of words containing unvoiced and voiced stops, we see that while equal percentages of high- and low-frequency words contain one or more unvoiced stops ($z = .47$), twice as many low-frequency words contain one or more voiced stops as do the high-frequency words ($z = 2.19$, $p < .025$). If we include the nasals n and m, respectively, in the groups of low- and high-complexity phonemes, we find again that equal percentages of high- and low-frequency words contain at least one

TABLE 4

PROPORTIONS OF LOW- AND HIGH-FREQUENCY STIMULUS WORDS WHICH CONTAIN AT LEAST ONE PHONEME FROM A GIVEN SET

Phoneme set	High-frequency words		Low-frequency words		Significance test for $P_H - P_L$
	Frequency	% (P_H)	Frequency	% (P_L)	
t, p, k (unvoiced)	14	38.9	16	44.4	$z = .47$
d, b, g (voiced)	9	25.0	18	50.0	$z = 2.19$*
t, p, k, n (low complexity)	20	55.5	21	58.3	$z = .24$
d, b, g, m (high complexity)	11	30.6	22	61.1	$z = 2.60$**

* $p < .025$.
** $p < .005$.

low-complexity phoneme ($z = .24$) and that twice as many low-frequency words contain at least one high-complexity phoneme as do high-frequency words ($z = 2.60$, $p < .005$).

While these results bearing on the "phonetic hypothesis" must be considered preliminary, they do point to the need for a systematic investigation of the phonetic correlates of word frequency, perhaps using different techniques for sampling representative groups of high- and low-frequency words. It is not known, at present, the extent to which the above results are dependent on the particular constraints used in selecting the samples of high- and low-frequency stimuli employed in the present study. In any event, it is clear that with disyllabic materials, *S*s, for whatever reason, do tend to form hypotheses which are in the same frequency class as the stimulus word, and this tendency introduces systematic differences into the *S*s' auditory recognition performance as a result of criterion bias in their rejection of previously held hypotheses.

Finally, the existence of phonetic features which distinguish words of high frequency from words of low frequency raises the additional possibility that such phonetic differences can themselves account for the word frequency effect. Suppose that those phonemes correctly perceived by *S* at a given S/N determine a response hierarchy composed primarily of words belonging in the same frequency class as that of the stimulus word. Zipf's law would then predict that the size of such a hierarchy would be smaller when the stimulus word was of high frequency than it would be if the stimulus word were of low frequency. This result suggests the possibility that on a simple (unsophisticated) guessing basis, high-frequency stimulus words will still be more often correctly guessed than will low-frequency words. The ability of such a mechanism to account for the magnitude of the word-frequency effect of course remains questionable, even if it can be established in the future that the phonetic differences between high- and low-frequency words are substantial.

REFERENCES

BROADBENT, D. E. Word-frequency effect and response bias. *Psychological Review,* 1967, **74,** 1–15.

BROWN, C. R., & RUBENSTEIN, H. Test of response bias explanation of word-frequency effect. *Science,* 1961, **133,** 280–281.

CATLIN, J. On the word-frequency effect. *Psychological Review,* 1969, **76,** 504–506.

FREDERIKSEN, J. R. Cognitive factors in the recognition of ambiguous auditory and visual stimuli. *Journal of Personality and Social Psychology Monograph,* 1967, **7** (1, Whole No. 639).

FREDERIKSEN, J. R. Response perseveration in auditory word recognition. *Journal of Experimental Psychology,* 1969, **79,** 48–55.

HAWKINS, J. E., JR., & STEVENS, S. S. The masking of pure tones and of speech by white noise. *Journal of the Acoustical Society of America,* 1950, **22,** 6–13.

HOWES, D. H. On the interpretation of word frequency as a variable affecting speed of recognition. *Journal of Experimental Psychology,* 1954, **48,** 106–112.

HOWES, D. H., & SOLOMON, R. L. Visual duration threshold as a function of word probability. *Journal of Experimental Psychology,* 1951, **41,** 401–410.

MILLER, G. A. The masking of speech. *Psychological Bulletin,* 1947, **44,** 105–129.

MORTON, J. A retest of the response-bias explanation of the word-frequency effect. *British Journal of Mathematical and Statistical Psychology*, 1968, **21**, 21–33.

POLLACK, I., RUBENSTEIN, H., & DECKER, L. Analysis of incorrect responses to an unknown message set. *Journal of the Acoustical Society of America*, 1960, **32**, 454–457.

SAVIN, H. B. Word-frequency effect and errors in the perception of speech. *Journal of the Acoustical Society of America*, 1963, **35**, 200–206.

SOLOMON, R. L., & POSTMAN, L. Frequency o usage as a determinant of visual duration thresholds of words. *Journal of Experimental Psychology*, 1952, **43**, 195–201.

THORNDIKE, E. L., & LORGE, I. *The teacher's wor book of 30,000 words*. New York: Columbi University, Teacher's College, Bureau of Publications, 1944.

ZIPF, G. K. *The psycho-biology of language*. Boston: Houghton-Mifflin, 1935.

(Received December 22, 1970)

Psychological Review
1971, Vol. 78, No. 5, 420–425

ON THE WORD FREQUENCY EFFECT: COMMENTS ON THE PAPERS BY J. CATLIN AND L. H. NAKATANI[1]

MICHEL TREISMAN[2]

University of Reading, England

D. E. Broadbent considered four models for the word frequency effect. He presented evidence rejecting the "sophisticated guessing" model and supporting a response bias model based on signal detection theory. J. Catlin and L. H. Nakatani have both proposed that Broadbent's arguments are faulty and that the "sophisticated guessing" account should be maintained. It is here suggested that it is their arguments that are at fault, and Broadbent's that stand.

The tendency for common words to be detected at lower signal-to-noise ratios than rare words, or to be detected more often at the same signal-to-noise ratio, is known as the word frequency effect. Broadbent (1967) discussed four possible explanations for this, labeling them the "pure guessing," "sophisticated guessing," "observing response," and "criterion bias" models. Green and Birdsall (1958) analyzed word recognition in terms of a signal detection model in which each word has its own decision axis. Broadbent's fourth model is similar to this, but in developing it he employed Luce's (1959) choice theory approach to provide an approximation to signal detection theory. He presented evidence which rejected the first three models and favored the last one. Recently, however, both Catlin (1969) and Nakatani (1970) have proposed that his reasoning is fallacious, and they have each concluded that the sophisticated guessing theory should be reinstated. The present note critically examines their arguments.

The sophisticated guessing model (Newbigging, 1961; Savin, 1963; Solomon & Postman, 1952) proposes that when a stimulus word is below its intelligibility threshold, some of its component features may nevertheless be detectable and may serve to delimit a small class of words consistent with them. Subjects' guesses are not then drawn at random from their whole vocabulary, but come from the subset of words so delimited, and the more common words are more likely to serve as guesses. Thus when the stimulus word presented in a recognition task is not detected, it is more probable that it will be guessed correctly if it is common than rare, with the result that frequent words will get higher correct detection scores. To illustrate this, let us assume a somewhat simplified situation in which presentation of the stimulus word results either in its detection, with probability $P(d)$, or in the delimitation of a set of words consisting of N_{H} high frequency words and N_{L} low frequency words, one word being selected from this set as a guess. For simplicity, assume that each high-frequency word has the same probability of appearing as a guess, and that the same is true for the low-frequency words, but that the probability of any one high-frequency word being guessed is x times $(x > 1)$ greater than the probability of any one low-frequency word being guessed. The probability of guessing an individual high-frequency word, when it has not been detected, is then $P(g) = x/(xN_{\mathrm{H}} + N_{\mathrm{L}})$; for a low-frequency word, it is $P(g) = 1/(xN_{\mathrm{H}} + N_{\mathrm{L}})$. When a stimulus word is presented, the probability that it will be correctly reported is then $P(r) = P(d) + [1 - P(d)]P(g)$, and the value of x will determine the magnitude of the word frequency effect.

The "error ratio" shall be defined as the ratio of the number of errors in which a high-frequency word is reported to the number of errors consisting of a low-frequency word when a given stimulus word has been presented for detection. The terms "high-frequency error ratio" and "low-frequency error ratio" will be used for the error ratios given when, respectively, high-frequency or low-frequency stimulus words are presented. Broadbent (1967) derived from the sophisticated guessing model the prediction that the high-frequency error ratio should be less than the low-frequency error ratio. If his arguments are applied to the present simplified illustration, they imply that when a high-frequency stimulus word is presented, the proportion of the responses which are high-frequency errors will be $[1 - P(d)] \times (N_{\mathrm{H}} - 1)x/(xN_{\mathrm{H}} + N_{\mathrm{L}})$ and the low-frequency error proportion will be $[1 - P(d)] \times N_{\mathrm{L}}/(xN_{\mathrm{H}} + N_{\mathrm{L}})$, giving a high-frequency error ratio of $x(N_{\mathrm{H}} - 1)/N_{\mathrm{L}}$. But when a low-frequency stimulus word is presented, the error ratio will be $xN_{\mathrm{H}}/(N_{\mathrm{L}} - 1)$, which will be greater. Moreover, to account for the magnitude of the word frequency effect, the difference between the two error ratios must be substantial. To illustrate this for the present example, let us suppose that $P(r|\mathrm{HF})$, the probability of correctly reporting a high-frequency stimulus word, is .33, and $P(r|\mathrm{LF}) = .13$, where $\mathrm{HF} =$ high frequency and $\mathrm{LF} =$ low frequency (these particular recognition probabilities were obtained by Broadbent, 1967, in an experiment in which subjects attempted to identify monosyllables presented in noise). Then $P(r|\mathrm{HF}) - P(r|\mathrm{LF}) = .33 - .13 = [1 - P(d)]x/(xN_{\mathrm{H}} + N_{\mathrm{L}}) - [1 - P(d)]/(xN_{\mathrm{H}} + N_{\mathrm{L}})$, which gives $x(5 - N_{\mathrm{H}}) \geqslant 5 + N_{\mathrm{L}}$ from which it follows that $N_{\mathrm{H}} < 5$. This would imply a considerable difference between the two error ratios.

---

[1] This paper was written while the author was in receipt of a grant from the Medical Research Council.

[2] Requests for reprints should be sent to Michel Treisman, Department of Psychology, University of Reading, Building 3, Earley Gate, Whiteknights, Reading, RG6 2AL, England.

If the sophisticated guessing model is correct, it appears, on Broadbent's interpretation of it, that the low-frequency error ratio should be found to exceed the high-frequency error ratio, but his experiments failed to demonstrate such a difference and Broadbent therefore rejected this theory. Catlin (1969), however, claims that the arguments which led him to predict unequal error ratios are wrong. His objection proceeds as follows. He starts with the assumption that on the average, the acoustical characteristics of high- and low-frequency words do not differ, that is, partial information will delimit similar acoustical neighborhoods in each case:

> Now, given that words of similar frequency . . . show no greater tendency to share acoustic features than do words of dissimilar frequency, the frequency distribution of words (excepting the stimulus word itself) in neighborhoods delineated by low-frequency words should be the same as the frequency distribution of words in neighborhoods delineated by high-frequency words [Catlin, 1969, p. 505].

Using an urn model, there will be S balls for the stimulus word, E_{H} balls for high-frequency potential errors contained in its acoustic neighborhood, and E_{L} balls for the low-frequency words in the neighborhood. The error ratio, $E_{\mathrm{H}}/E_{\mathrm{L}}$, will then be independent of the frequency of the stimulus word, and Broadbent's experimental data, therefore, are completely consistent with sophisticated guessing theory.

Although the suggestion that the acoustical properties of words are unrelated to their frequency of usage has not been demonstrated to be correct, and quite possibly does not hold exactly, it is probably an acceptable approximation, and has in practice been incorporated into sophisticated guessing theory. It underlies both Broadbent's and Catlin's arguments. Their difference lies in the implications they take to follow from it. Broadbent's argument has been discussed. Catlin claims that "the principle of acoustical equivalence implies that the frequency distribution of words *other than the stimulus word* in the acoustical neighborhood of a stimulus word will be independent of stimulus-word frequency [p. 505]." However, Catlin does not demonstrate that the exclusion of the stimulus word from the relevant acoustical neighborhood can be derived from the assumption that the frequency and acoustical characteristics of words are independent; he treats this as a self-evident consequence. It is important, therefore, to examine whether it does in fact follow from this assumption.

In approaching this question, it will be useful to consider a hypothetical dialect of English which contains only the following monosyllables ending in "-at": high-frequency words—cat, hat, pat, fat; low-frequency words—tat, vat. For each high-frequency word, we suppose $P(g) = .2$; for each low-frequency word, .1 (i.e., we take $x = 2$). It is then natural to suppose that if a word is presented and not recognized, but the partial cues "monosyllable" and "terminal -at" are detected, then these will delimit the whole of the "-at subset" above, whether these partial cues have in fact come from the high-frequency stimulus word "cat" or from the low-frequency stimulus word "vat." Since these partial cues exclude any other word in the language from occurring as a response, it follows that if "cat" is presented but not correctly identified, the error ratio will be $.6/.2 = 3$, but when "vat" is presented, it will be $.8/.1 = 8$. It is not possible to envisage any operation on these partial cues which could realize Catlin's expectation that high- and low-frequency error ratios should be the same unless the assumption that the subject has detected only cues common to both the possible stimulus words under consideration is violated. This assumption requires that the word emitted as a response will depend on the structure of the subset alone, and not in any way on the actual stimulus word presented. When the response happens to coincide with the stimulus word, this will not be counted as an error, and thus the different error ratios given above necessarily follow. Obviously the same sort of result would be true of any other "common" subset, that is, any subset delimited by cues common to both the stimulus words being considered, such as, in this instance, "- -t" or "-a-."

We now need to examine whether this conclusion will also hold for "specific" sets, that is, subsets delimited by partial cues peculiar to one or the other stimulus word, but not to both. When "cat" is presented and not fully identified, the partial cues detected might be "monosyllable" and "ca-" and these might determine the specific subset: high-frequency words—cat, can; low-frequency words—cad, cam. When "vat" is presented, detection of "va-" might delimit: high-frequency words—vac, van; low-frequency words—vat, valve. If we assume that for each high-frequency word in these subsets $P(g) = .333$, and for each low-frequency word, .167 (i.e., we again take $x = 2$), it follows that when "cat" is presented and "ca-" detected, the resulting high-frequency error ratio is $.33/.33 = 1.0$; when "vat" is presented and "va-" detected, the low-frequency error ratio is $.66/.167 = 4$. Thus a difference in the error ratios is present and is in the same direction as before when the subsets specific to each stimulus word are considered. It is, of course, obvious that we get this result because we have assumed that the two specific subsets have essentially the same structure. We could avoid this result by assuming that the "ca-" subset is constituted by high-frequency words—cat, can, cap, cab, catch, each with $P(g) = .167$; low-frequency words—cad, cam, each with $P(g) = .083$; it would then follow that when "cat" is presented, and this subset delimited, the resulting high-frequency error ratio would be $.66/.167 = 4$, the same ratio as for the "va-" subset. Evidently, by increasing the number of high-frequency words in the "ca-" subset still further, we could produce an even larger high-frequency error ratio, and the point could be reached where the advantage to the overall high-frequency error ratio derived from trials on which specific subsets of this sort are activated could just compensate for the larger low-frequency error ratio given when common subsets are delimited, and thus Catlin's prediction would at last be achieved. But this success would have its cost. If this final assumption is taken to illustrate a feature of the language, and not just of the comparison of "cat" and "vat," then, in general, partial cues from high-frequency stimulus words must tend to delimit subsets which have a higher proportion of high-frequency words in them than is the case when the partial cues come from low-frequency stimulus words. Thus, the two populations of partial cues must be different: to achieve acoustical neighborhoods which are the same *when the stimulus word is excepted* in the sense that they give Catlin's prediction, the essence of the "principle of acoustical equivalence" has had to go by the board.

It might seem that if it will keep sophisticated guessing theory afloat, this ballast, though precious, could happily be sacrificed. But in fact, not even this will save Catlin's prediction comfortably. For it is not sufficient to jettison acoustical equivalence. The difference in the specific subsets that would be required to equalize the high- and low-frequency error ratios when the stimulus words we choose to compare are "cat" (high) and "vat" (low) will have the opposite effect, worsening the disparity between the error ratios, when it is "van" (high) and "cam" (low) that come under examination. The number of ad hoc assumptions that would be required to balance out this and all other such contrasting effects for the language as a whole would clearly be so large, and so arbitrary, that even if this is possible, it would obviously be quite unacceptable. Thus it seems fair to conclude that Catlin's defense of sophisticated guessing theory must fall away.

Nakatani (1970) attacks Broadbent's distinction between the "pure guessing" and the signal detectability "criterion bias" models. He claims to demonstrate that the two are "mathematically equivalent"; this rests on a proof that the models "both predict the same set of confusion matrices." If this demonstration is valid, its implications are far-reaching; they require to be considered carefully. Nakatani proves this "equivalence" for 4 × 4 confusion matrices, but as he notes, his proof can be generalized to any confusion matrix. In assessing what weight should be given to his argument, it will be convenient to apply it to the 2 × 2 probability matrix for the standard yes–no signal detection task. In terms of what Swets (1961) has called "high-threshold theory," this is predicted by the following matrix:

$$M_1 = \begin{array}{c} \\ \text{SN} \\ \text{N} \end{array} \begin{array}{c} \begin{array}{cc} \text{Yes} & \text{No} \end{array} \\ \begin{bmatrix} p+(1-p)g & (1-p)(1-g) \\ g & (1-g) \end{bmatrix} \end{array} \qquad [1]$$

M_1 gives the probabilities of the responses Yes and No when the subject is presented with signal plus noise (SN) or with noise (N) alone. The true probability of detecting the signal is given by p; the corresponding probability for noise is effectively zero, and there is a guessing rate parameter, g. Using Luce's (1959) "choice theory" parameters to give an approximation to the signal detectability model, we get:

$$M_2 = \begin{array}{c} \\ \text{SN} \\ \text{N} \end{array} \begin{array}{c} \begin{array}{cc} \text{Yes} & \text{No} \end{array} \\ \begin{bmatrix} \dfrac{\alpha}{\alpha + v} & \dfrac{v}{\alpha + v} \\ \dfrac{1}{1 + v} & \dfrac{v}{1 + v} \end{bmatrix} \end{array} \qquad [2]$$

where α and v are parameters relating to the separation between SN and N means, and the criterion level, respectively (Broadbent, 1967; Luce, 1959). The pure guessing model is essentially a generalization of high-threshold theory, and Broadbent uses Luce's parameters in developing his criterion bias model.

In this case, Nakatani's argument should proceed as follows. By applying the substitutions

$$\begin{aligned} v &= (1 - g)/g, \\ \alpha &= \{p + g(1 - p)\}/g(1 - p), \end{aligned} \qquad [3]$$

or their converses, any matrix in the set predicted by high-threshold theory can be shown to be contained also in the set predicted by the response bias model, and vice versa. This renders the models "formally equivalent."

From this application of Nakatani's argument, it is evident that if it is correct to accept that Broadbent's distinction between the pure guessing model and his "apparently new" criterion bias model is "misleading" because the models are "formally identical" (Nakatani, 1968), then we are obliged also to accept that Tanner and Swets' (1954) distinction between threshold theory and signal detection theory is equally vacuous. This will not be generally acceptable, nor is it difficult to find reasons for questioning this argument. First, we note that in the yes–no case, each matrix represents a possible data point on or above the positive diagonal in the unit square. What Nakatani's equivalence establishes, and all that it establishes, is that any point which can be fitted by the one model, in the sense that that point could be derived from allowable parameter values, can also be fitted by the other. That is, each model can accommodate the same range of potential data. However, being applicable to the same data is not a sufficient criterion for establishing the identity of alternative models. Indeed, on this argument, any further theory, for example, Luce's (1963) version of threshold theory, which is capable of placing a receiver operating characteristic curve at any specified point on or above the positive diagonal would also be "formally identical" with the two models of the yes–no task discussed above.

It is generally accepted that models which generate the same predictions can be regarded as equivalent. But the minimum requirement for this conclusion is not that a given prediction, or class of predictions, should be identical, but that all possible predictions of both models should be the same. In the present case, the models are clearly capable of producing contrasting consequences. Thus, the receiver operating characteristic curve, which can be derived from M_1 by eliminating g, is linear, whereas M_2 gives a nonlinear curve symmetric about the negative diagonal. Similar differences can be derived from the pure guessing and criterion bias word frequency effect models.

Nakatani presents a further argument for rejecting the criterion bias model. Applied to matrix M_2 it would run as follows. If $P(\text{No}|\text{SN}) = v/(\alpha + v) = 0$, then we must have $v = 0$. But then we should get $P(\text{No}|\text{N}) = v/(1 + v) = 0$. However, data may give $P(\text{No}|\text{SN}) = 0$, but $P(\text{No}|\text{N}) > 0$. Thus, the signal detectability model (or the choice theory approximation to it) is not reasonable and cannot be maintained. Extended to larger matrices, the argument is that if one cell in a column has a zero value, so should all. Since confusion matrix data that contradict this prediction are available, the criterion bias model "is clearly contrary to intuition and data" and must be rejected.

Unfortunately this rests on an assumed infallibility of data. It does not allow for random error. In practice, it is never possible to say with complete certainty that a zero response rate represents a true zero probability, rather than a small nonzero value. For this reason, it has never been suggested that experimental findings of the form $P(\text{No}|\text{SN}) = 0$, $P(\text{No}|\text{N}) > 0$, require the rejection of signal detectability theory (or its "choice theory" approximation). For the same reason, this sort of finding constitutes unimpressive evidence against the criterion bias model for word recognition.

Finally, both Catlin (1969) and Nakatani (1970) emphasize that the criterion bias model imposes no restriction on the words which may manifest as errors, whereas sophisticated guessing theory suggests, and Savin (1963) has found, that errors tend to be acoustically similar to the stimulus word. Newbigging (1961) has reported similar findings for visual stimuli. However, these findings pose no great difficulty for a signal detection theory of word recognition. It is entirely possible to suppose that when partial cues compatible with a limited number of words are received from a stimulus, the perceptual system will produce a rise in the mean value of the central distribution on the decision axis corresponding to each of those words. Erroneous responses would then manifest the joint effects of the lowered criteria applying to high-frequency words and the rise in the means of the corresponding central distributions suffered by words acoustically similar to the stimulus word, and this would explain the clustering of errors found by Savin (1963) and others.

REFERENCES

BROADBENT, D. E. Word-frequency effect and response bias. *Psychological Review*, 1967, 74, 1–15.

CATLIN, J. On the word-frequency effect. *Psychological Review*, 1969, 76, 504–506.

GREEN, D. M., & BIRDSALL, T. G. *The effect of vocabulary size on articulation score.* (Tech. Memo No. 81.) University of Michigan, Electronic Defense Group, 1958.

LUCE, R. D. *Individual choice behavior.* New York: Wiley, 1959.

LUCE, R. D. A threshold theory for simple detection experiments. *Psychological Review*, 1963, **70**, 61–79.

NAKATANI, L. H. *A confusion-choice stimulus recognition model applied to word recognition.* (Tech. Rep. No. 31) University of California, Los Angeles, Department of Psychology, Human Communication Laboratory, 1968.

NAKATANI, L. H. Comments on Broadbent's response bias model for stimulus recognition. *Psychological Review*, 1970, **77**, 574–576.

NEWBIGGING, P. L. The perceptual redintegration of frequent and infrequent words. *Canadian Journal of Psychology*, 1961, **15**, 123–132.

SAVIN, H. B. Word-frequency effect and errors in the perception of speech. *Journal of the Acoustical Society of America*, 1963, **35**, 200–206.

SOLOMON, R. L., & POSTMAN, L. Frequency of usage as a determinant of recognition thresholds for words. *Journal of Experimental Psychology*, 1952, **43**, 195–201.

SWETS, J. A. Is there a sensory threshold? *Science*, 1961, **134**, 168–177.

TANNER, W. P., & SWETS, J. A. A decision-making theory of visual detection. *Psychological Review*, 1954, **61**, 401–409.

(Received March 12, 1971)

Psychological Review
1971, Vol. 78, No. 5, 426–450

ISSUES IN PSYCHOPHYSICAL MEASUREMENT[1]

S. S. STEVENS[2]

Harvard University

Two classes of ratio-scaling procedures are outlined—magnitude matching and ratio matching—and their assets and liabilities are noted. Partition-scaling procedures, which are supposedly designed for interval scaling, produce results that can be described by a power function with a virtual or "as if" exponent. Since the virtual exponent is smaller than the actual exponent of the continuum, the category scale is nonlinear. The virtual exponent provides a convenient descriptor of several kinds of partition operations. Other topics discussed include individual differences among subjects' exponents, procedures of averaging, and the effects of stimulus range on exponents. It is suggested that the power law asserts a nomothetic imperative.

The task here is to review the matching procedures used to determine the power functions that govern the growth of sensation magnitude and to consider some of the sources of deviation and perturbation that have raised questions concerning the nomothetic quality of the psychophysical power law.

Since all procedures of measurement involve matching operations, the interesting differences among different scales and different kinds of measurement can often be reduced to a basic question: What was matched to what, and how? In the domain of psychophysics, numerous scaling methods have been invented, many of them useful for the determination of ratio scales of apparent magnitude. The approaches of ratio scaling can be catalogued in different ways, but for present purposes they fall into two general classes: *magnitude matching*, which includes the subclasses (*a*) cross-modality matching, (*b*) magnitude estimation, and (*c*) magnitude production; and *ratio matching* which includes the subclasses (*a*) cross-modal ratio matching, (*b*) ratio estimation, and (*c*) ratio production.

Since there are endless variations on psychophysical procedures, it is possible here to comment on only a few of the features that characterize the principal methods. Consideration is also given to some of the interval or partition methods, and to the distinction between the virtual exponent and the actual exponent. Other problems discussed concern individual differences, averaging, and range effects.

MAGNITUDE MATCHING

These procedures include all direct equations between two continua. Three principal varieties of magnitude matching have been distinguished.

Cross-Modality Matching

When the stimulus for a continuum can be readily varied by means of a control of some kind, it becomes possible to match that continuum to any other continuum. Figure 1 gives examples of matching functions produced when several different continua were matched to vibration on the fingertip.

Ideally, the experiment comparing two continua should provide for a balanced design in which each continuum is matched in turn to the other continuum. A balanced procedure may help to assess and correct the regression effects that are always present in the matching operation (Stevens & Greenbaum, 1966). In the typical experiment, the observer tends to shorten the range of whichever variable he controls. Even within the same modality the regression effect shows up in matching functions. Thus, two somewhat

---

[1] Supported in part by Grant NS-02974 from the National Institutes of Health (Laboratory of Psychophysics Report ppr-366-133).

[2] Requests for reprints should be addressed to S. S. Stevens, Laboratory of Psychophysics, Harvard University, 33 Kirkland Street, Cambridge, Massachusetts 02138.

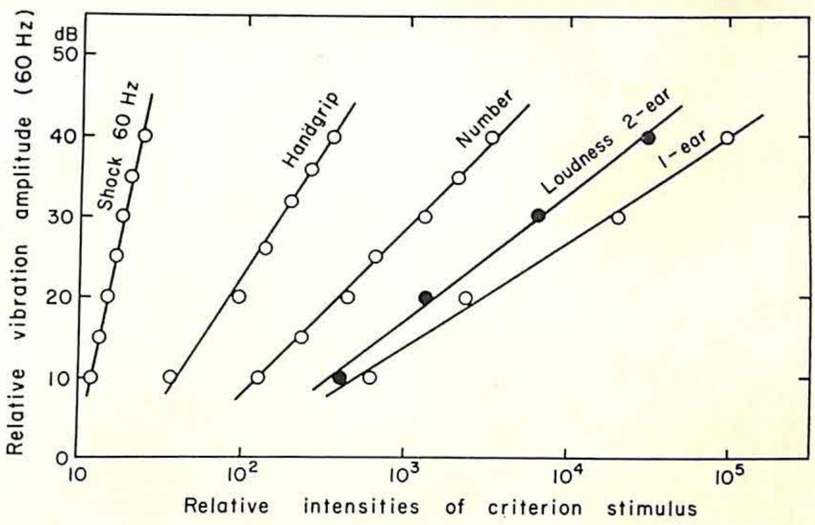


FIG. 1. Equal sensation functions obtained by cross-modality matches between various continua and a 60-Hz. vibration on the fingertip. (The vibration amplitudes were set by the experimenter. The observer adjusted the other stimulus to produce an apparent match—from Stevens, 1968a.)

different functions were obtained, depending on which auditory stimulus the subject adjusted in matching a tone to a noise.

Because it is often difficult to give the subject control of the stimulus, many cross-modality comparisons have not yet been made. It is difficult, for example, for the subject to vary the heaviness of lifted weights in order to match heaviness to loudness.

When a balanced design in the matching of two continua is impracticable, an evaluation of the regression effect may sometimes prove possible by way of a third continuum. Two principal paradigms can be distinguished.

1. Continuum A is adjusted to match each of two continua, B and C. The ratio of the exponents of the matching function A to B and A to C determines the exponent of the function relating B to C. The derived function would presumably be free of the regression effect, provided the regression occasioned by adjusting A remained constant when the criterion continuum was changed from B to C. The hoped-for constancy might be upset by such factors as disparities in difficulty or range.

2. A second procedure for counterbalancing the regression effect can be utilized whenever it is possible to match both A and B to a common continuum C. The two matching functions provide exponents whose ratio determines the exponent of the power function relating A to B. Whether the regression effects in the two matching functions A to C and B to C are exactly equal may not be known, of course, but the procedure may still cancel a major portion of the regression bias.

An instance of the second paradigm was provided by Moskowitz (1969), who asked observers to match both numbers and loudness to a wide variety of taste mixtures. From each of the 68 pairs of experiments that Moskowitz conducted we can derive an estimate of the power-function exponent relating number to loudness. The geometric mean of the 68 estimates was .67, which agrees with the value of the exponent adopted for loudness calculation (Stevens, in press). The standard deviation was .51 decilog, or about 12%. That degree of scatter may be regarded as an empirical guide to the amount of variability to be expected in tests of transitivity among the exponents of the power functions. When the regression effect has been canceled, about two-thirds of the measured exponents may be expected to lie within plus or minus half a decilog of the predicted exponents (see Stevens, 1969).

The regression effect, of course, is only one of the sources contributing to the systematic errors that affect the outcome of experiments, but it is one of the most obstinate, and therefore perhaps the most important. And it may be composed of more than a single factor.

Magnitude Estimation

This procedure is actually a form of cross-modality matching in which numbers are matched to stimuli. When first used, the procedure was called absolute judgment (Stevens, 1953), later numerical estimation (Stevens, 1954), and still later magnitude estimation (Stevens, 1955a). That last name appears to have stuck. In this context, the number continuum can be regarded as another perceptual modality. Magnitude estimation or "number matching" has become a popular method, mainly because of its convenience. The subject brings the numbers with him, so to speak, and the experimenter needs only to provide the target stimuli to which the numbers are to be matched. The nature of the task can be portrayed in terms of a typical set of written instructions to the subject.

> You will be presented with a series of stimuli in irregular order. Your task is to tell how — they seem by assigning numbers to them. Call the first stimulus any number that seems to you appropriate. Then assign successive numbers in such a way that they reflect your subjective impression. For example, if a stimulus seems 20 times as —, assign a number 20 times as large as the first. If it seems one-fifth as —, assign a number one-fifth as large, and so forth. Use fractions, whole numbers, or decimals, but make each assignment proportional to the — as you perceive it.

Experience has shown that it is usually better not to designate a standard. The subject then remains free to choose his own modulus. If possible, stimuli should be presented in a different irregular order to each subject, but the first stimulus is usually chosen from among those in the middle region, rather than from one end of the range. Between 10 and 20 stimuli may be presented at a session. A good schedule provides for one judgment, or at the most two judgments, per stimulus by each subject. After the subject has learned to recognize a particular stimulus, little or no new information is obtained from subsequent judgments of its repeated presentation. Furthermore, biases due to range and spacing of stimuli seem to have less effect when the subject is limited to one judgment per stimulus.

Untrained, inexperienced college subjects seem to do as well at the matching tasks as those who have had many years of practice. Hence, there is no need to "train" the subjects. Indeed, since there is no right or wrong to the subjects' responses, it is not clear what would be meant by training. Under some circumstances, the nature of the task may profitably be clarified by allowing the subjects to begin by matching numbers to an easier continuum, such as apparent length of lines, or apparent size of circles.

Averaging can be done by computing geometric means or medians. The log-log slope (exponent) determined by the geometric means is not affected by the fact that each observer uses a different unit of modulus. When it is desired to adjust the judgments to a common modulus, a good method is to minimize the squares of the individual subject's intercept differences. The procedure is: convert all scores to logs, compute grand mean of logs, and adjust each log score for each observer by whatever additive constant makes the observer's mean correspond to the grand mean. That procedure of *modulus equalization* permits each of an observer's estimates to contribute to the correction factor to be applied to that observer's modulus.

Magnitude Production

Here the experimenter presents the numbers one at a time in irregular order, and the subject adjusts the stimulus to produce an apparent match. The numbers themselves should normally approximate a geometrical progression. For example, in an extensive study of loudness and its inverse, softness, the successive numbers presented were in the ratio 2 to 1 and ranged from 1.25 to 80 (Stevens & Guirao, 1962). Sample results are shown by the triangles in Figure 2.

Because of the regression effect, the power functions obtained by magnitude production are typically steeper (have larger exponents) than those obtained by magnitude estimation. An unusually large regression angle is illustrated in Figure 2. It is often assumed that the unbiased function lies between the two functions obtained by estimation and production, and indeed it may. But where? Are the error sources in the one procedure exactly balanced by the error sources in the other? Although exact balance may be possible, it seems hardly likely that no asymmetry exists. An average function may nevertheless be desired, in which case it may be well to compute the geometric mean of the two exponents. The geometric mean is invariant under an interchange of the two coordinates (Indow & Stevens, 1966). When the results of magnitude estimation and production are combined in an appropriate way, the combined procedure offers advantages over either procedure alone.

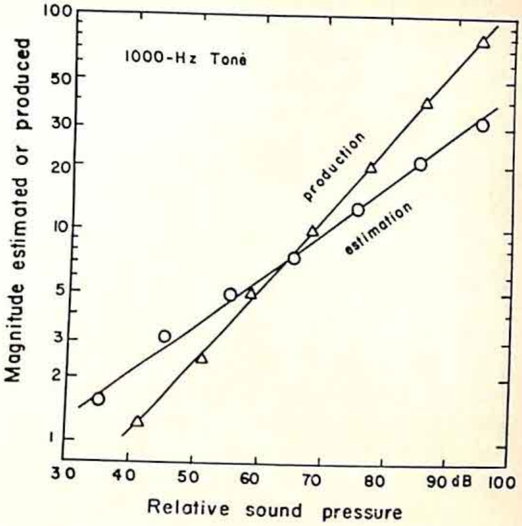


FIG. 2. Magnitude estimation and magnitude production of loudness. (Each point is the geometric mean of two estimates or two productions by each of 10 observers.) (Reprinted with permission from an article by J. C. Stevens and M. Guirao published in the *Journal of the Acoustical Society of America*, 1964, Vol. 36. Copyrighted by the Acoustical Society of America, 1964.)

A particular version of the combined procedure designed to produce a balanced function has been spelled out by Hellman and Zwislocki (1968) and applied to loudness functions.

Another way of combining some of the features of production and estimation is to permit the subject to set stimulus levels at his own pleasure and to report the apparent magnitude. The experimenter in effect gives up all control over the stimuli. A radical procedure of that kind was used by J. C. Stevens and Guirao (1964) to show that individual subjects produce power functions and that the power function is not, as some writers had suggested, an artifact of averaging. The key parts of the instructions were as follows:

> Your task is to set the tone to different levels of loudness and to assign numbers to each of the loudnesses. Make your numbers proportional to the loudness you hear. You may make as many settings as you want. Try to cover a wide range of loudness.

Eleven subjects, chosen at random from among students, staff, and secretaries, gave the results shown in Figure 3. One subject made as few as seven production estimations. Another made five times that many. Two subjects made settings that extended over almost 100 decibels (db.) and made estimations that ranged over about 10,000 to 1, which implies an exponent near .8. The experiment was repeated some months later. The geometric mean of the 22 exponents for the 11 subjects was .7, which is fairly close to the value 2/3, which has been proposed as the standard value (Stevens, in press).

RATIO MATCHING

The earliest form of ratio matching appears to have been devised by Merkel (1888) in order to determine what he called the "doubled stimulus." It was a direct ratio-scaling method, but its potentialities were not effectively exploited. The method would now be classed as one of the varieties of ratio production. It is convenient to distinguish three subclasses of ratio matching, as follows:

Cross-Modal Ratio Matching

J. C. Stevens set two different brightnesses in front of the subject and asked him to adjust one of two noises to make the ratio of the noises match the apparent ratio of the brightnesses (see S. S. Stevens, 1961a, pp.

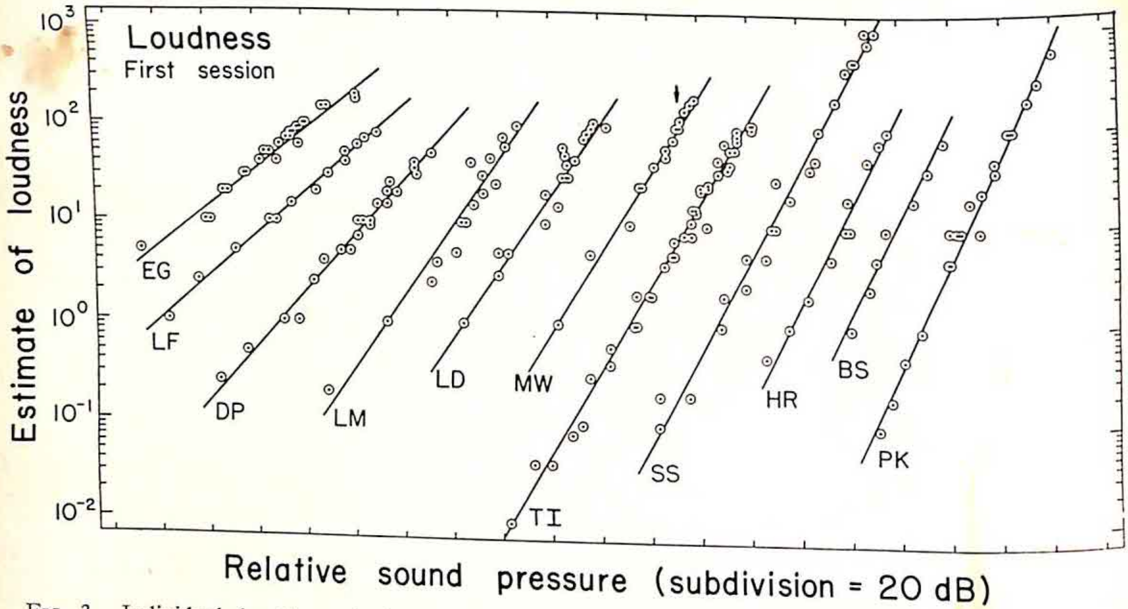


FIG. 3. Individual functions obtained when each of the 11 observers set the stimulus level and estimated the loudness. (Each point represents a judgment. There was no averaging—from J. C. Stevens & Guirao, 1964.)

17–18). The results confirmed the relative values of the exponents for loudness and brightness, showing, in fact, that the two exponents are approximately equal.

Ratio Estimation

Here the subject matches numerical ratios to apparent stimulus ratios. In the "complete" version of the procedure, stimuli are presented in all possible pairs and the apparent ratios are estimated (Ekman, 1958). Other versions use fewer stimulus pairings, and some versions provide for reporting in terms of fractions or percentages. For example, in an early experiment, Ham and Parkinson (1932) presented a sound at one level followed by a sound at a lower level, and asked the subjects to estimate what percentage of the loudness remained.

Ratio Production

In this once-popular scaling procedure, the subject is required to find or produce the stimulus that seems to stand in a prescribed relation to a standard stimulus. As we have seen, Merkel invented that kind of task with his method of doubled stimulus. Fractionation is the name commonly used for procedures that require the subject to set a stimulus to one-half (or some other fraction) of the standard. (For a tabulation of many of the numerous ratio productions that have been made with acoustic stimuli, see Stevens, 1955a.)

Ratio production has fallen into disuse mainly because magnitude matching seems to be a superior procedure. The biases in ratio production are such that the method often fails to produce a clean power function.

INTERVAL MATCHING AND VIRTUAL EXPONENTS

Although the judgment of intervals or differences, as required in various kinds of partitioning operations, may produce satisfactory results on metathetic continua, systematic biases afflict partitions carried out on prothetic continua. Furthermore, the partitioning operations can produce at best an interval scale, not a ratio scale. Nevertheless, one or another form of interval matching has produced data that have played a role in the establishment of the psychophysical power law (Stevens, 1953).

In order to describe the procedures used and the results obtained in the various kinds of partition operations, it is convenient to distinguish two exponents: a virtual or functional exponent and the actual exponent of the continuum in question. The virtual exponent is the one the observer appears to be using when he makes his partition judgments. It is an "as if" exponent. The value of the virtual exponent α turns out to be lower than that of the actual exponent β. Since $\alpha < \beta$, the scales created by partitioning are nonlinear relative to the corresponding magnitude scales created by magnitude or ratio matching.

Perhaps the best known example of a partition scale is the Munsell scale for the lightness of grays. That scale has been determined and redetermined by several kinds of partition operations. A series of gray papers may also be scaled by magnitude estimation, as was shown by Stevens and Galanter (1957). The virtual exponent of the Munsell scale is approximately .33 and is decidedly lower than the actual exponent, approximately 1.2, obtained by magnitude estimation.

Let us now consider four varieties of interval scaling procedures.

Cross-Modal Interval Matching

This procedure seemed to work well in a 1953 experiment when subjects adjusted markers along a line (position, a metathetic continuum) in order to match the apparent spacing of a series of loudnesses. Subjects also matched marker position to the apparent spacing of the heaviness in a series of lifted weights. The same principle is involved, of course, in numerous rating scales: the subject expresses his opinion by marking a position on a line. Newhall (1950) used a somewhat similar method, involving markers on a two-dimensional grid, in order to determine spacings among the apparent lightnesses of gray papers. His results agreed with the Munsell scale. As a general method, however, interval matching suffers from a basic ambiguity, especially when metathetic position is not one of the continua used. If two prothetic continua are involved, the subject may find it easier to match ratios than differences, and without intending it, he may actually produce ratio matches between the pairs of stimuli on the two continua.

Efforts to judge intervals often encounter a dramatic hysteresis effect, which makes the judgment highly contingent on the order in which the stimuli are presented (Stevens, 1957b).

Interval Estimation

Here the subject may be asked to assign numbers to represent the sizes of apparent differences. For example, Dawson (1968) presented pairs of loudnesses and asked the observers to make a magnitude estimation of the apparent difference in each pair. He also presented pairs of visual areas. The typical biases that emerge under partitioning procedures were apparent in the results, especially in the judgments of loudness differences. A constant loudness difference (in sones) is not judged to be constant; rather a given difference is judged smaller when it is moved up the stimulus scale.

Another demonstration of the bias in interval judgments—the operation of a virtual exponent—is contained in the results of Beck and Shaw (1967) who asked 28 subjects to judge four loudness intervals, 5, 10, 15, and 20 sones in width, each located at four stimulus levels. The median estimations for three of the interval sizes are shown in Figure 4 as a function of the sound pressure level of the tone at the lower end of the interval. The curved lines show the general trend of the data.

If the world were so constructed that equal prothetic intervals appeared equal to the perceiving subject, the lines in Figure 4 would be straight and horizontal. The downward trend of the data in Figure 4 illustrates the typical result obtained in partition judgments of whatever variety: equal intervals are not judged equal at different locations on a prothetic continuum. As an interval of a constant size moves up the scale of the continuum, the constant interval is judged to be smaller and smaller.

The curves in Figure 4 were generated by a partition model in which it was postulated that the observer's judgments are governed

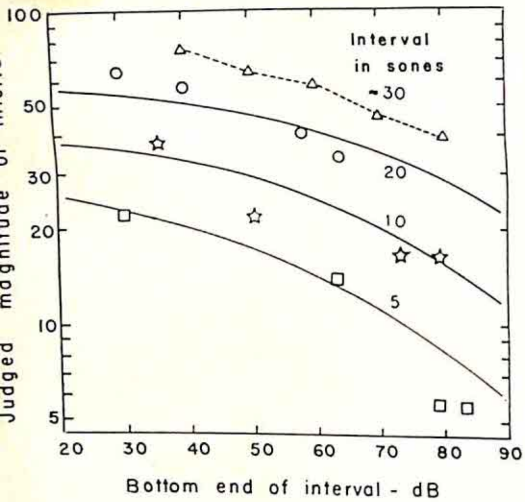


FIG. 4. Showing how the judgment of an interval of a constant size depends on the location of the interval. (Observers made magnitude estimations of sets of intervals 5, 10, 20 sones wide. In another set of experiments (triangles) the intervals were approximately 30 sones wide. The stimulus level at the bottom end of the interval is shown by the abscissa. The ordinate gives relative values only. As a constant interval moves upward in sound pressure level, the perceived size of the interval decreases. The family of three curves was generated by assuming that instead of the actual exponent of the sone scale .6, the observers used a virtual or "as if" exponent equal to .3. Triangles from Dawson, 1968; other data from Beck & Shaw, 1967.)

by a power law that does not have the actual exponent of the continuum, but rather has a virtual or functional exponent equal to .3. The fit of the curves is only fair, for the data do not provide enough information to distinguish between the family of functions generated by the virtual exponent value .3 and the family given by some other nearby exponent. If the observer's virtual exponent were .6, it would correspond to the actual exponent, and the lines in Figure 4 would then become straight and horizontal. As the virtual exponent becomes smaller, the family of curves tilts more steeply downward, and the distance between the curves decreases.

To a first approximation, then, it appears that the observer judges loudness intervals as if his power function had a virtual exponent about half as large as the actual exponent of the continuum. That principle was rather nicely confirmed in a second experiment by Beck and Shaw (1967) in which 29 observers made magnitude estimations of another set of loudness intervals—intervals that were constructed to be constant in size as determined by the lambda scale, a scale that was constructed so as to agree with a particular set of bisection data (Garner, 1954). Over the stimulus range of interest here, the lambda scale has an effective exponent of approximately .3. In other words, the "constant" intervals provided by the experimenters were generated by a function whose exponent coincided with that of the virtual exponent. As we should expect, therefore, the judged size of the intervals did not show a downward drift with increasing stimulus level. In fact, when the judgments are plotted as in Figure 4, but with *lambda interval* rather than *sone interval* as the parameter, the data describe functions that are very nearly horizontal. Thus the principle is clear: When the generating function used to set up the equal intervals has the same exponent as the virtual operating function employed by the observers in their partition judgments, then the intervals all appear equal.

It is of interest next to consider the other extreme and to ask what happens when the equal intervals are generated by a function with an exponent that is lower than the virtual exponent. Since a power function with a very low exponent resembles a logarithmic function (see Figure 5), we can examine the problem by setting up equal logarithmic or equal decibel intervals. Again the results turn out as expected. Decibel intervals appear to grow larger as their absolute level is raised. Thus a series of successive 10-db. intervals beginning at 40 db. produced the following magnitude estimations: 1.51, 1.89, 2.65, 4.14, and 9.10 (Dawson, 1968). Those values represent averages over four different experiments. They show that the 10-db. interval 80–90 db. was judged to be about six times larger than the 10-db. interval 40–50 db. If plotted in Figure 4, the judgments of 10-db. intervals would describe a curve that sweeps upward rather than downward. In other words, observers' judgments demonstrate that the virtual exponent is decidedly greater than zero, and that a logarithmic function does not accord with interval judgments.

As a matter of fact, although Dawson did not use intervals that were exactly constant in sones, his data agree fairly well with functions having the form of those generated by a virtual exponent of about .3. The triangles in Figure 4 show the magnitude estimations of the largest sone intervals employed, ranging from 16 to 31 sones. The magnitude estimations of those intervals were multiplied by a factor in order to bring each of them to the value it would have had if all the intervals had been 30 sones. That multiplicative correction appears to do little or no violence to the data. The path depicted by the triangles in Figure 4 shows how the apparent size of a 30-sone interval grows smaller as the stimulus level used to define the bottom end of the interval increases.

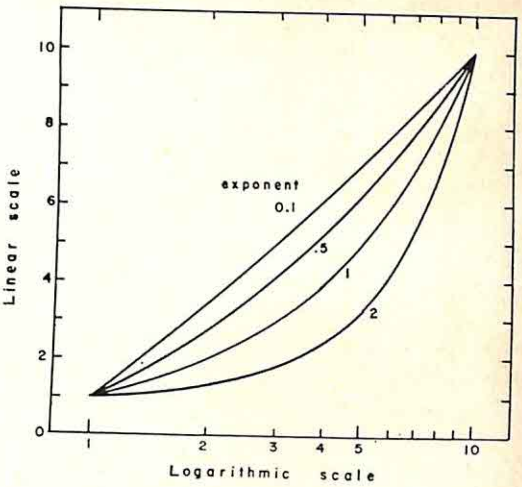


FIG. 5. Sample power functions plotted in semilogarithmic coordinates. (As the exponent of the power function decreases, the graph of the function becomes straighter. It thereby approaches the form of a logarithmic function, which is described by a straight line in these coordinates.)

Interval Production

The production of prescribed intervals was the method invented by Plateau (1872). He asked eight artists to paint a gray such that the intervals from gray to black and gray to white would appear equal. That type of procedure is often called bisection. When more than two equal intervals are involved, it is called equisection (Garner, 1954).

The intervals to be produced need not be equal. For example, in one of my own experiments, observers produced loudness intervals corresponding to markers spaced unevenly along a line. The results from that procedure of multisection were consistent with the results obtained by equisection.

Under conditions designed to make the judgment maximally easy, the results of bisection experiments have sometimes been found to agree fairly closely with the psychophysical functions produced by magnitude and ratio matching. In other words, although the inevitable systematic difference was clearly evident, the size of the difference proved to be fairly small (see Stevens, 1955a for loudness, 1961b for brightness). To a limited extent, then, the bisection experiments have confirmed the exponents of the magnitude scale obtained by ratio-scaling procedures. The magnitude-scale exponents serve as an upper bound and the bisection-scale exponents normally lie below the bound.

In a bisection experiment in which ϕ_2 is set midway between ϕ_3 and ϕ_1 the virtual exponent of the power function can be determined by the bisection equation (Stevens, 1955a), which may be written

$$\phi_3{}^{\alpha} - \phi_2{}^{\alpha} = \phi_2{}^{\alpha} - \phi_1{}^{\alpha}.$$

The exponent α may be found by iteration. It seems to be an invariant rule that, as with other partitioning procedures, the value of the virtual exponent determined from bisection has a lower value than the exponent determined by procedures that call for ratio judgments.

The family of curves in Figure 4 tells us that a true bisection on a prothetic continuum will not appear correct to the observer. The lower half of the interval will appear larger than the upper half. Consequently, when the observer himself makes the bisection, he lowers the bisecting point. In some bisection experiments, the bisection point has fallen low enough to agree with the exponent used to generate the curves in Figure 4 (Garner, 1954). In other experiments, the virtual exponent has been found to lie closer to the actual exponent (Stevens, 1955a).

Sum Production

The virtual exponent describes the convexity in the partition function under still another procedure. The bisection problem can be turned around, so to speak, and instead of asking the observer to produce segments, the experimenter can present the segments and ask the observer to produce the whole, or the sum. If the virtual function is convex, we should expect that the perceived whole would prove to be less than the sum of the perceived parts. When observers were shown two or more line segments and asked to produce a line that appeared equal to the sum, the line produced was longer than the sum of the separate lengths (Krueger, 1970). In other words, a line equal to the sum of the separate lengths would have appeared shorter than the apparent sum. The experimental results demonstrate that the exponent 1.0, which observers ordinarily use in judging apparent length, has been replaced by a virtual exponent having a lower value.

RELATION TO POWER LAW

Perhaps the most important outcome of the work on partition scales lies in the demonstration, by means of interval methods, that the psychophysical function is a power law. The argument rests on the fact that the apparent bisection point remains approximately constant when all the stimuli are increased or decreased by the same factor. For example, Plateau's eight artists each worked in a different atelier, under a different level of illumination, and yet they all produced bisecting grays that, when viewed in the same setting, were *"presques identiques."* When that happens, the sensory magnitude must follow either a logarithmic law or a power law. If the bisection point falls at the geometric mean between the end points used to define the bisected interval, then the log function is indicated. If the bisection point falls above the geometric mean, as it seems to have done in Plateau's experiment, then a power function is indicated. And the virtual exponent α increases as a function of the distance of the bisecting point above the geometric mean. Some of those basic relations are illustrated in Figure 6.

Under optimal experimental conditions, the virtual exponent α, determined by bisection, may approach the value of the actual exponent of the continuum as an upper bound. Normally, as we have seen, the bisection exponent lies below the continuum exponent. Nevertheless, the invariance of the bisection exponent under multiplicative variations in the stimulus levels has provided evidence that equal stimulus ratios produce equal sensation ratios—which is the invariance principle that underlies the power law.

CATEGORY SCALES AND THRESHOLDS

The category scale calls for special consideration because it is by far the most common and yet perhaps the least satisfactory form of partition scale. Most of the many users of category scales throughout science, education, engineering, and commerce do not intend a partitioning operation. Their hope is to grade or assess some variable, but they proceed to prescribe—and to limit!—the subjects' response scale. Commerce does much of its buying and selling with the aid of the grading of goods by subjective assessment on simple category scales, a crude procedure perhaps, but it serves its limited purpose. The business of scaling and measurement is not a primary aim in industry as it is in psychophysics. Nevertheless, category scaling is occasionally used in psychophysical experiments, despite the demonstrated inefficacy of the procedure. Stevens and Galanter (1957) examined some 70 different category scales on a dozen different perceptual continua and found little to justify the use of category scaling in quantitative studies. Their hope was that category scaling would thereafter fall into disuse. They hoped in vain.

The categories may be designated by a limited set of adjectives, such as large, medium, and small. Or the categories may be designated by a finite set of numbers, such as 1 to 6. Those were the numbers used for history's first recorded category scale, the scale of stellar magnitude, which dates from about 150 B.C. and which in a much revised form still serves the astronomer (see Stevens, 1960).

On prothetic continua, the category scale is invariably nonlinear relative to the magnitude scale. When plotted against the scale obtained by magnitude estimation, for example, the data from category judgments produce a curve that is concave downward. Whenever the subject is asked to categorize, he is forced to divide the continuum into parts or segments in order to make it conform to the limited, finite set of numbers or adjectives that he is required to use. In other words, he is obliged to attend to differences or distances. Under those circumstances, the subject is forced out of ratioing and into partitioning.

What we learn from category experiments is that the human being, despite his great versatility, has a limited capacity to effect linear partitions on prothetic continua. He may do quite well, to be sure, if the continuum happens to be metathetic, but, since most scaling problems involve prothetic continua, it seems that category and other forms of partition scaling ought generally to be avoided for the purposes of scaling. If, for some reason, an unbiased interval scale is needed, it can be obtained from a ratio scale, for the ratio scale contains the interval scale (Stevens, 1946).

The reverse is not possible, however. The ratio scale cannot be recovered from the interval scale when only interval information is available.

Is there a use for category scaling? Although essentially useless for ratio scaling, category methods play an indispensable role in threshold measurements, where the problem reduces to the determination of boundaries between classes. In fact, all threshold determinations involve category procedures, because the problem is to sort stimuli into classes, for example, those that are detectable and those that are not, those that are disturbing and those that are not, or those that are acceptable and those that are not, and so on.

Psychophysical thresholds are boundaries between classes. Although the boundary that we call a sensory threshold may be sharp at a given point in time, in a living organism the boundary behaves as though it were jittering about. Consequently, the measurement of a threshold becomes a statistical process: On the basis of repeated samples, we make a statistical decision regarding the location of the boundary. Still, the underlying experimental operation for determining any kind of threshold always involves a procedure of matching either stimulus to category or category to stimulus.

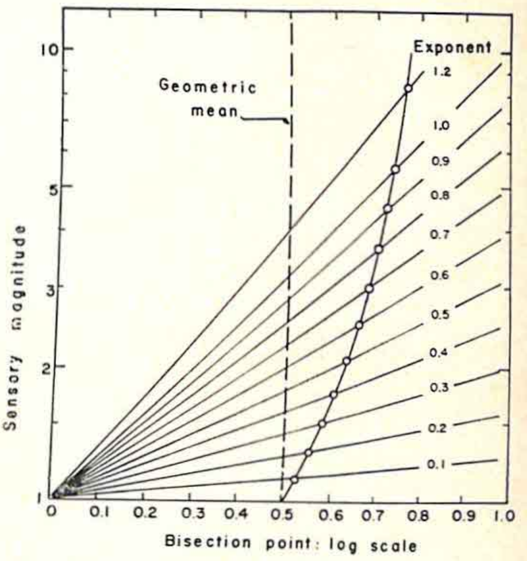


FIG. 6. Schematic diagram showing a family of power functions with exponents ranging from .1 to 1.2. (As shown by the circles, the position of the bisection point moves upward along the stimulus scale (abscissa) as the exponent increases. Bisection at the geometric mean would correspond to a logarithmic function, or a vanishingly small exponent.)

POWER-GROUP TRANSFORMATIONS

As we have seen, partition scales can often be usefully described by power functions. Consequently, the nonlinearity of the partition scale can be conveniently described by the difference between the virtual exponent of the partition scale and the actual exponent of the magnitude ratio scale. In other words, the operation of partitioning causes the observer to behave as though a power-group transformation had been performed, that is, as though an exponent had been altered. Whenever two power functions differ in exponent, they are nonlinearly related. Since the virtual exponent is always less than the actual exponent, the curvature of the partiloudness, measured in sones, grows very large as loudness increases.

In order to illustrate how unlikely, indeed impossible, it would be for an observer to make a correct judgment of a loudness difference regardless of stimulus level, let us consider the following paradigm. We will assume that the loudness exponent is .6 and that the interval from 40 to 50 db. corresponds to the subjective interval between a loudness of 1 sone and a loudness of 2 sones, or a difference of 1 sone. That happens to be a clear and obvious difference. Now consider the same 1-sone difference at 100 db. where the loudness is 64 sones. An increase from 100 to 100.2 db. makes a 1-sone difference, raising the loudness from 64 to 65 sones. But a change as small as .2 db. is so small that it would be detected less than half the time. It seems beyond possibility, therefore, that a constant difference of 1 sone could be judged to be the same size regardless of where it occurred on the continuum.

That illustrative example may help to suggest why it is that every variety of partitioning must give a distorted result on a prothetic continuum. A constant difference transforms itself from obvious to undetectable as we go from weak to strong stimuli. Therein lies the essence of the prothetic principle.

A graphic illustration of the underlying principle is shown in Figure 7. The ordinate shows the stimulus change in decibels that corresponds to a change of 1 sone. As in the example above, at a level of 40 db. a change of 1 sone corresponds to 10 db. At 100 db., the required change is only .2 db., and it becomes still smaller at higher levels. The jnd (Weber fraction) is plotted in Figure 7 as a horizontal line with an ordinate value of about .5 db. (see Miller, 1947). A plot similar to Figure 7 can be constructed for any prothetic continuum. For visually presented lengths, for example, such a plot could be made to show how, in visual perception, a centimeter added to a centimeter makes a clearly perceived difference, whereas a centimeter added to a meter becomes only marginally detectable. Under successive visual presentation, length behaves like loudness, and a constant added difference becomes lost to view as the stimulus increases.

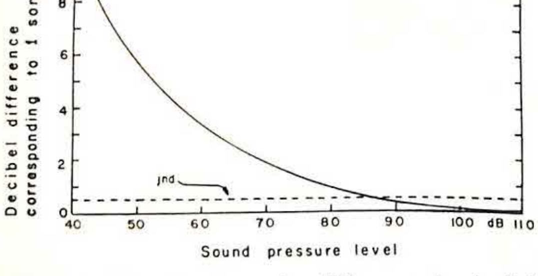


FIG. 7. Showing how the difference in decibels required to produce a 1-sone loudness difference falls off with increasing sound pressure level. (At 40 db. it requires an added 10 db. to add 1 sone. At 110 db. it requires an added .1 db. to add 1 sone. The dashed line shows the approximate value of the jnd, the increase that would be detected about half the time.)

The smooth continuity of the curve in Figure 7 suggests that the underlying process that forces a given constant difference to become less and less apparent as the level increases is a process that operates all up and down the prothetic continuum. There are no discontinuities, no sudden transitions. When the observer attempts a comparison of differences at any place on the continuum, it is as though his perception undergoes an asymmetrical distortion; a constant difference seems larger toward the lower than toward the higher part of the continuum. From that basic asymmetry, it follows that the operations of partitioning on prothetic continua will fail to produce a linear, unbiased interval scale. Where the asymmetry does not exist, as on a metathetic continuum, it becomes possible for partitioning to produce a linear interval scale.

INDIVIDUAL DIFFERENCES

Let us turn at this point to a problem that was especially well formulated by Jones and Marcus (1961). Do different individuals have different operating characteristics in their sensory systems? If not, then why do the slopes (exponents) differ when observers match numbers to stimuli, as, for example, in Figure 3?

If we were to take some of the measured values literally, it would be incumbent on us to examine the implications. Consider in Figure 3 the functions for observers EG (exponent .4) and PK (exponent 1.1). They both happen to be "normal" observers, having similar audiometric thresholds. If, as seems reasonable, the two observers experience the same loudness when a tone is near threshold, what about a tone 100 db. above threshold? The two exponents suggest that the loudness at 100 db. would be about 3,000 times greater for PK than EG. Yet both observers react very similarly to acoustic stimuli. For example, both observers report that 100 db. sounds very loud; they both call 120 db. slightly painful; and they both jump when it is first turned on. Both observers set the intensity control of the radio to roughly the same level for comfortable listening. Both carry on conversations in about the same level of voice. All those bits of evidence make it hard to credit a difference in exponent that entails a 3,000-fold difference in loudness at 100 db.

Another consideration is this. The value of an exponent reflects the curvature in the operating characteristic of the sensory system. As a function of sound pressure, the exponent .4 suggests that the loudness function for EG is a decelerating function, sharply concave downward. The exponent 1.1 suggests that the function for PK is concave upward—an accelerating function of sound pressure. In order to achieve such an unlikely difference in curvature, nature would have had to produce two radically different kinds of operating characteristics in the two auditory systems.

An alternative hypothesis is that all human auditory systems operate on much the same design, with very nearly the same exponents, but that there are wide individual differences in what observers take to be the numerical value of a loudness ratio. We encounter that kind of individual difference especially clearly when an observer is asked to adjust one sound to make it appear half as loud, or twice as loud, as another sound. In an experiment with 22 observers, the scatter of eight distributions of settings covered a median range of 14 db. (For plots of the distributions, see Stevens, 1957a.) That range of settings would correspond to a range of exponents from about .3 to 1.5. The distributions of the ratio settings tended to be roughly log normal, however, and there appeared to be no reason not to average the data, which is equivalent to averaging individual exponents.

Consider this question: What would happen if an experimenter were to report that each of 22 observers in a group made an identical setting when asked to produce a two-to-one loudness ratio? It is an interesting question whether the scientific community would regard such a result as a new marvel, or whether scientists would simply reject it as an implausible outcome. We expect variability and we usually find it. The substantive question concerns whether and how to average the data.

Although the operating characteristics of normal sensory systems may have the same or closely similar exponents, there are circumstances in which an abnormal exponent can be demonstrated and in which the abnormality calls for careful measurement. The otologist, contemplating middle-ear surgery, needs to know whether so-called recruitment, with its attendant large exponent, characterizes the lower part of the patient's loudness function. If the exponent is abnormally large, middle-ear surgery is contraindicated. But if the exponent is normal, the otologist may feel free to try to correct a middle-ear difficulty. (For a model depicting the manner in which Ménière's disease may produce power-group transformations on the loudness function, see Stevens & Guirao, 1967.)

Except for clinical cases, it seems fair to say that seldom, if ever, have investigators undertaken the serious effort that is needed to establish the existence of individual differences in sensory power functions. Mostly, we are shown the variability that happens to be found in a particular matching response, usually magnitude estimation. The obvious next step, usually not undertaken, would be to try magnitude production. I have found that many individual differences vanish as soon as the matching task is inverted. For example, by magnitude estimation, the lowest loudness exponent in a group was .4, but by magnitude production, that particular observer produced the exponent .9; the geometric mean of the two exponents is .6 (see graphs in Stevens & Greenbaum, 1966).

Five observers in those same experiments (Stevens & Guirao, 1962) made both magnitude estimations and magnitude productions of the loudness of noise. The ratio of the largest to the smallest exponent for estimation was 1.42; for production, it was 1.27. When the geometric means of the exponents for estimation and production were examined, the range ratio fell to 1.14. In other words, the individual differences were less pronounced when both estimation and production were used in a balanced design.

The balancing of estimation by production may be good for a start, but if we are seriously interested in the power function for a particular individual, we will not stop with magnitude estimation and production. We will want to know the results from a balanced array of additional cross-modality matching tasks, the more the better.

Variability and Averaging

Criticism of the power law has sometimes centered on one or another aspect of variability. It would be good, of course, if variability could be reduced, so that the psychophysical functions could be determined with higher precision. But empirical functions always suffer from variability, and the central question is not so much whether a measurement is variable, or whether subjects disagree, but whether averaging is appropriate. If the data can be appropriately averaged, it does not matter how widely the variability may range, provided the number of independent measurements can be increased. In principle, the standard error can then be brought down to any desired level.

In electrophysiology, for example, miracles of averaging are performed routinely by computers programmed to dig a particular waveform out of the myriad variations in the ongoing neural activity of the brain. A repeated click delivered to the ear can then be seen as an evoked potential at the scalp. The response to the click emerges remarkably clear and unencumbered by the noise of the brain, for the noise has been suppressed by being averaged out. An analogous strategy of error cancellation by averaging finds usefulness in psychophysics and in all the rest of science.

In the long run, since scientists tend to believe only those results that they can reproduce, there appears to be no better option than to await the outcome of replications. It is probably fair to say that statistical tests of significance, as they are so often miscalled, have never convinced a scientist of anything. By contrast, a tabulation of 178 determinations of the loudness exponent, based on 25 years of accumulated results from several different laboratories, produced a median result .6, which became the exponent recommended by the International Standards Organization (see Stevens, 1955a). Since then, a further accumulation of experimental determinations has begun to fix the second decimal place, and it now appears that the value 2/3 may be more representative (Stevens, in press). The value 2/3 happened to correspond to the modal value of the 1955 distribution, but at that time, the median seemed a better choice than the mode.

How to average data presents serious and interesting questions. The median is perhaps the single most unbiased measure of location, and it has often been used in psychophysics. A popular rule is: when in doubt, use the median. On the other hand, a more efficient average is often wanted, and the choice of an efficient measure can usually be made to rest on the form of the distribution. Thus, two different averages have proved appropriate in psychophysical scaling, each under a particular circumstance.

In an equisection experiment, 45 subjects divided a 40-db. segment of the loudness continuum into four equal-appearing intervals. Because the decibel measures of the subjects' settings gave skewed distributions, it did not seem proper to average the decibel values, which would have been equivalent to computing geometric means. When the decibel measures were converted into sones (a linear loudness value), the settings showed the desired symmetries. It was concluded that averaging should be done by computing the arithmetic means of the loudness values, and an iteration procedure for determining those values was outlined (Stevens, 1955b). The arithmetic mean is also an appropriate average for other kinds of partition scales, such as the category scale.

In experiments involving magnitude estimation and other forms of cross-modality matching, it is the geometric mean, not the arithmetic mean, that appears to be the appropriate average. Against a linear scale, the response distributions are skewed, as indeed they must be when error is proportional to magnitude. When error grows in proportion to magnitude, so that the relative error stays constant, a logarithmic transformation tends to undo the skewness. The applicable principle states that when error is relative, the error distribution is log normal.

The log-normal model for magnitude estimation was tested by J. C. Stevens who plotted the results obtained when 70 naive observers estimated the loudness of a white noise presented by a loudspeaker in a classroom (J. C. Stevens & Tulving, 1957). In the first of two experiments, the observer chose his own modulus by assigning to the first stimulus (85 db.) whatever number seemed appropriate. In the second experiment, the stimuli were presented in pairs, a standard at 85 db. called 10 followed by a variable. When the cumulative frequencies for this second experiment were plotted on probability paper against the logarithm of the magnitude estimations, the result was a family of straight lines. In other words, the distributions were log normal.

Some five years later J. C. Stevens and Miguelina Guirao applied the modulus equalization procedure to the data of the first experiment, the one in which each of the 70 subjects had chosen his own modulus. (The deviation of each subject's scores from the average function was minimized by the procedure for modulus equalization described above.) The cumulative frequencies of the judgments subjected to modulus equalization are shown in Figure 8. Again the straight lines demonstrate that the distributions are approximately log normal.

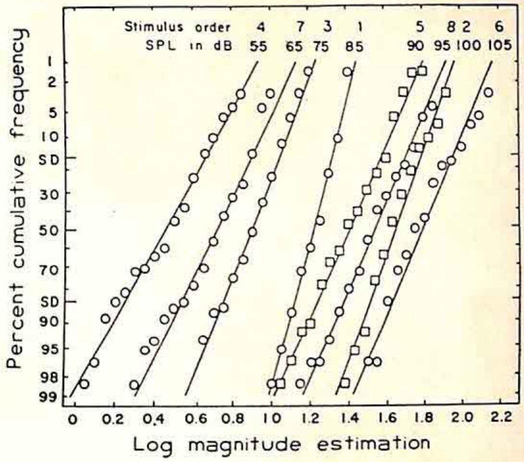


FIG. 8. Cumulative frequency distributions of magnitude estimations. (Each of 70 subjects, making their first judgments ever, assigned whatever number seemed appropriate to eight levels of white noise presented in the order shown. The data of three subjects who used negative numbers or zeros were not tabulated. In these coordinates, a straight line signifies a log-normal distribution. Data from J. C. Stevens & Tulving, 1957.)

Other features of the results in Figure 8 are also of interest, especially in view of the fact that those were the first magnitude estimations ever made by that large group of naive listeners. Since it was a classroom experiment, the order of the stimuli could not be made different for the different listeners. Consequently, the order of the stimuli is reflected in the slopes of the cumulative frequency lines, or, in other words, in the standard deviations. The first stimulus has the smallest standard deviation (1.0 decilog), the second stimulus the next smallest (1.4 decilog), and so forth. If the stimulus order had been made different for each observer, the lines in Figure 8 would be more nearly parallel. But a tendency has been observed in numerous experiments for the variability to increase at the low end of the scale, and to a lesser extent, at the high end. That same tendency is apparent in Figure 8. Hence, it can be seen that the effect due to stimulus order cuts across the normal tendency for the variability to be slightly lower in the middle range.

The lines in Figure 8 make it clear that the geometric mean is an appropriate average for the data. In this instance, the geometric means determine a power function with the exponent .55.

A similar treatment was applied to the data from the original "no standard" experi-

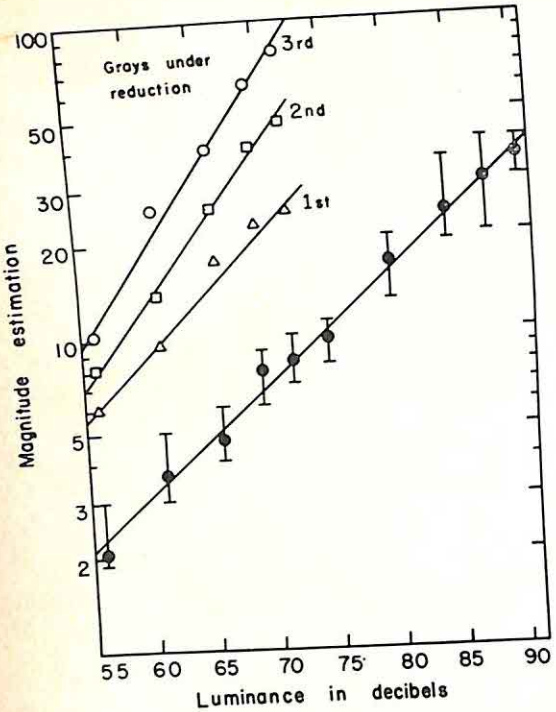


Fig. 9. Showing how Munsell grays, extending from black to white, are judged when they are presented entirely alone with no other visible luminance in the field of view. (The luminance then covers a range of about 16 db., which seems subjectively rather short. On the three successive presentations of the set of stimuli, the exponent became larger (unfilled symbols). The filled symbols show the results for an extended luminance range. The exponent is .35, which is much lower than the value 1.2 obtained when the Munsell grays are placed one by one on a table in front of the observer.)

Counterbalancing for Range

The general problem of the interaction of range and exponent may be thought of in terms of the matching of two continua, A and B, each of which may serve in turn as the adjusted and as the criterion continuum. What does the observer tend to do to the adjusted stimulus as a function of the range (logarithmic difference) between two criterion stimuli set by the experimenter? For purposes of discussion let us assume that the true exponent determined by the slope A/B in log-log coordinates is 1.0.

If the experimenter were to set the criterion range to a very small value, say, to a small fraction of a decilog, on the average the observers would necessarily respond with adjusted ranges that average larger than the criterion range. That value of the adjusted range would then determine an exponent greater than 1.0. On the other hand, if the criterion range was set at a very large value, the observers would tend to match it with shorter ranges, determining thereby an exponent smaller than 1.0. Between those two extremes, the exponent would decrease monotonically as the range increased.

Since the two continua A and B can be interchanged in their roles of criterion and adjusted variable, in a balanced experiment two paths would be traced by the exponent as the criterion range was varied from very small to very large. Those two paths would cross, and the crossing point would presumably determine the unbiased exponent—unbiased by the effect of range, but not necessarily free of other possible biases.

A concrete example of the two paths followed by the exponent when range is varied in a fairly well-balanced experiment can be constructed from the data of Stevens and Poulton (1956). The loudness function for a 1000-Hz. tone was studied by allowing groups of 8 to 11 unpracticed observers to make only a single judgment, either a ratio estimation or a ratio production. A standard stimulus was sounded first and called 100, and the observer expressed the apparent ratio by assigning a number to a second stimulus at a lower level. The exponents corresponding to the median estimations of each group are shown by the triangles in Figure 10. Each triangle represents a different group of listeners.

For the ratio productions, the observers adjusted a sone potentiometer to produce a prescribed fractional loudness relative to a standard. The exponent corresponding to the ratio productions (sone average) for each group of subjects is shown by a circle in Figure 10. Again each circle represents a different group of listeners. The data show an approximate symmetry and suggest that for those naive observers making their first loudness judgments, the exponent lies between .6 and .7. The range effect emerges as a dramatic but orderly variable, and since it shows an approximate symmetry, it can in

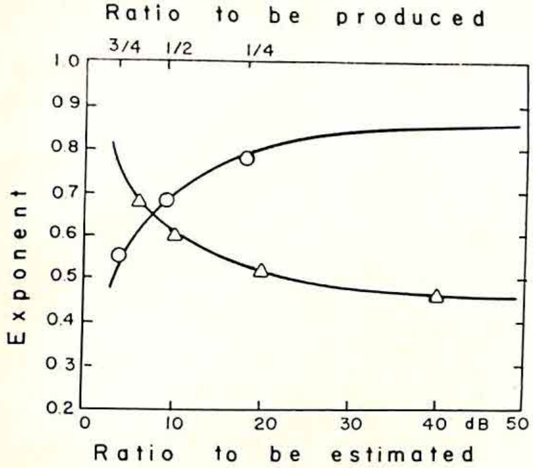


FIG. 10. The range effect in a partially balanced experiment. (Each point represents a separate group of subjects who estimated a loudness ratio (triangles) or produced a loudness ratio (circles). When the range effect produces symmetrical functions, as is approximately true here, the exponent may be uniquely determined. It is the value that makes the exponent corresponding to the crossover point equal to the exponent implied by the relation of the scales at the top and bottom of the figure. The exponent so determined is .65. Data from Stevens & Poulton, 1956.)

principle lead to a unique exponent determined by the crossing point in Figure 10.

The actual crossing point depends, of course, on the relation between the two scales (top and bottom in Figure 10) against which the two sets of data are plotted. But the relation between the values on the two scales (both logarithmic) also expresses an exponent. The problem, then, is to adjust the relation between the two scales (ratios to be produced and ratios to be estimated) so that the exponent determined by the scale relation coincides with the exponent determined by the crossing point. The adjustment can be carried out by iteration.

In plotting Figure 10, I first assumed that the exponent was .60. Accordingly, the ratio 1/2 at the top of the graph was set directly above 10 db. at the bottom of the graph. The other ratios were then set according to a logarithmic spacing. When the data were plotted, the crossover was found to correspond to .64 on the ordinate—a value that was larger than my assumed exponent. I next assumed a larger exponent, .67, and changed the upper scale and the location of the circles accordingly. The resulting crossover then fell at .65—a value smaller than my assumed exponent. Thus the two assumptions, one too high and one too low, had succeeded in bracketing the exponent between .64 and .65, but closer to .65. With fallible empirical data, the exact value of the exponent cannot be taken too literally, but it is nevertheless interesting that the crossover value accords approximately with the consensus of other measurements.

The search for other instances in which a pure range effect could be studied has been only partially successful. Experiments with the required balanced design seem to be rare, but we can compare two separate experiments that add up to a partly balanced design. In Figure 11, the triangles show the exponents corresponding to the median ratio estimations made by groups of about 30 subjects (Poulton, 1969). Each group judged one noise level set at 5, 20, or 35 db. above a standard noise (600 to 1200 Hz.) at 65 db. The circles in Figure 11 represent the exponents determined from the ratios produced

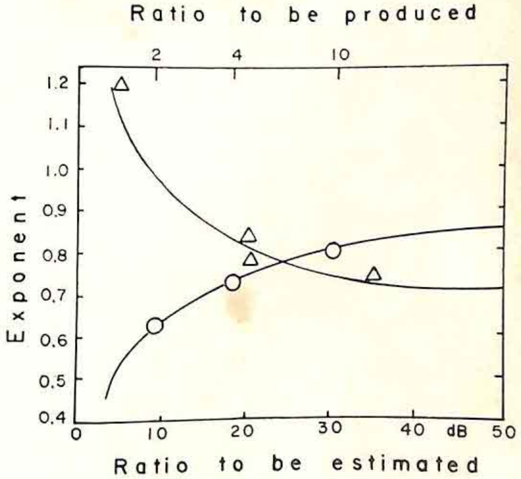


FIG. 11. Range effect shown in two noncomparable experiments. (Separate groups of subjects estimated loudness ratios (triangles) of a band of noise, 600–1200 Hz. (Poulton, 1969). A single group of subjects produced fractional and multiple ratios (circles) of a 40-tone complex (Geiger & Firestone, 1933).)

by 31 listeners who made multiple and fractional settings of noise consisting of a 40-tone complex (Geiger & Firestone, 1933). Both the multiple and the fractional productions were averaged to determine the exponents. The same group of listeners heard all the stimuli, and the order of the ratio productions was 1/2, 1/4, 1/10, . . . 2, 4, and 10.

In order to plot the circles in Figure 11, the exponent was assumed to be .67, and the ratio 2 on the upper scale was set directly above 9 db. on the lower scale. The crossover point then fell at .77. If the exponent is assumed to be .8, and if the scale values and the circles are moved accordingly, we find that the crossover point becomes about .78. Thus the two assumed exponents have bracketed the crossover exponents. Consequently, the exponents that would be determined by the data in Figure 11, if they were homogeneous and suitable for such a purpose, would lie between .77 and .78. That value is on the high side, even though the measured exponents for noise are often larger than those for tones. The mean exponent for white noise in an earlier compilation was approximately .72 (Stevens, 1955a).

The groups of subjects who gave the results that determined the triangles in Figure 11 subsequently judged all the other stimuli. The pooled results determine a power function with an exponent of about .73, provided no attempt is made to force the best fitting line to pass through the standard stimulus. Although experimenters have sometimes assumed that the function must pass through the standard, there is in fact no such requirement. The observer does not respond to any stimulus, whether standard or variable, with zero error.

The foregoing examples of the sources of error and distortion that may plague a psychophysical measurement are more illustrative than exhaustive. As has often been said, there are many ways to perform a bad experiment. Not even the expert in statistical design can tell exactly how to perform a good one. In principle, however, we can study each experimental ailment and utilize the rules of its behavior in order better to diagnose the nomothetic substrate.

THE NOMOTHETIC IMPERATIVE

The scientist's contest with nature has prospered to the degree that simplicities and uniformities have been detected amid the complexities that afflict observation and experiment. The simple invariances have often proved hard to find, however, because no experiment can be performed without its "context," and no measurement can be made without error. In psychophysics, each experimenter records results that disagree to some extent with those of his colleagues, and a penumbra of uncertainty surrounds even our best determinations. Consequently, there is room for many hypotheses and for many views regarding the structure of the psychophysical domain.

First there are those whose working hypothesis states that there exist laws to be discovered. Heeding the nomothetic imperative, those investigators refuse to be put off by the apparent chaos in the organism's reactions to stimuli, and they try to order, classify, and systematize the behavioral facts. Some people seem willing to gamble that the sense organs operate in beautifully simple ways and that what we take to be complexity lies more in our inept descriptions than in nature's actual comportment. As regards the particular question of reactions to stimulus intensity, the nomothetic outlook assumes that there exists an orderly input–output function, a simple law of some sort—however difficult it might be to pin it down in its exact form. Perhaps it was that same nomothetic outlook that drove Kepler to search for simple invariant laws in the baffling paths of the wandering planets, which were thought in early times to crisscross the skies impelled by their own volition.

The nomothetic viewpoint is consonant with an objective, operational approach, but it does not necessarily entail a particular philosophy of reality. In the fitting of schematics to empirical fact—what I have called the schemapiric endeavor—there is no necessity that we take a stand regarding any ultimate concern. We can simply note, for example, that when the intensity of the visual stimulus increases, the observer's response changes in an orderly fashion. If he is instructed to squeeze a hand dynamometer to match the apparent intensity of what he sees, then the force of his squeeze increases as a power function of the physical luminance (J. C. Stevens, Mack, & Stevens, 1960). That is the first-order fact. But what shall we say about it? We have obviously measured something, but have we measured sensation? As soon as that word appears, a discordant chatter sets in, because many authors contend that sensation is not a thing that can be measured.

If I thought it would help, I should happily give up the word sensation in favor of some other term, such as behavioral response or apparent effect, but the philosopher would shortly discover that my new term is only a euphemism for what I regard as a straightforward construct, a construct that can perhaps best be communicated to other people by my labeling it sensation.

A prime source of confusion in this semantic issue rests with the fact that each of us experiences sensations, and to each of us our sensations seem personal and private and inaccessible to measurement. As scientists, we should try to ignore that fact. To help free us of narcissistic introspection, there is the example of Plateau, the blind physicist, who gave us our first interval scale for the lightness of grays. He himself did not need to see the grays that the eight artists painted, for he could record the reports of other observers. Plateau's achievement drives home a crucial point: the blind could develop the psychophysics of vision; the deaf could develop psychoacoustics.

Sensation then is a construct, a name given to a constellation of behaviors. The justification for saying that sensation is subjective is that a human subject exhibits the behaviors. The heating coil in an electric stove also exhibits interesting behaviors, but quite arbitrarily we refrain from using the term subjective for those behaviors or for the constructs built on them. The simple, operational dichotomy into subjective (meaning people) and objective (meaning not people) seems eminently convenient, but words and slogans sometimes cause clash and conflict. In the schemapiric view of science, words and symbols serve only the neutral purpose of implementing a schematic structure which may be related by operational rules to an empirical structure (see Stevens, 1968b).

As an experimenter, I feel free to use terms like sensation and subjective, because they can be defined operationally in the context of psychophysical research. For some writers, however, the use of such words puts the user in the camp of Subjectivism, as opposed to Behaviorism (e.g., see Baird, 1970). I have long thought of myself as a behaviorist, but it has not seemed defensible to assert that verbal taboos provide a tool for advancing science, as some behaviorists have seemed to believe. A viable injunction is this: Use any words you care to, but let it be plain what operations lie behind your verbal forms. Otherwise said, construct the schema as you will, but tell us precisely how the empirics articulate with the schematic terms in order to produce the schemapiric substance.

Such protestations, I realize, will avail nothing with those who want to maintain a philosophical distinction between certain traditional views and to preserve a dualism that to an operationist is devoid of meaning. Since I do not believe in the usefulness of the distinction, I am happy to read that

> It is impossible to locate Stevens' view precisely, since he constantly shifts back and forth between a behaviorist and some other way of treating the question . . . [Savage, 1970, p. 390].

More damning than that, however, the whole psychophysical enterprise is said by Savage to be wrongly conceived, so that

> However we reconstrue Stevens' law, it cannot be construed as one relating sensations to stimuli, since the former are incapable of measurement [p. 541].

More than half a thousand pages, it should be said, are devoted by Savage to the thesis that we cannot measure what psychophysicists seem so delightedly to be measuring.

Philosophical problems must not detain us, however, for philosophical issues are eternal and do not go away. More tractable, perhaps, are some of the specific issues that strike directly at the nomothetic imperative. A cluster of questions has concerned the problem of picking the best sensory scale.

When three different scales have been created on the same continuum by three separate sets of operations—jnd's, category estimation, and magnitude matching—on what grounds do we presume to ascribe to one of those scales a superior position? Helson (1964) seemed to deny that we can make any such judgment, for he said that

> no one scale, however carefully established, can be considered better than other scales obtained under different conditions of judging [p. 179].

Other authors have also lamented the absence of criteria for determining the validity of a choice of one kind of scale over another. Validity is indeed the issue here. Which scale best measures what it is that we want to measure? Since it is that kind of question, the answer becomes a matter of opinion —a value judgment. It appears that all problems related to validity must seek their ultimate solution in the pragmatic domain. If a scale does the job we want done, we usually accept it.

But opinions differ about problems and solutions. How else can we understand the decision of an experimenter to limit the observer's responses to a finite set of numbers, such as 1 to 7 or 1 to 20? That curious maneuver of constraining the observer's responses is a tactic that seems somewhat mulish to those who have allowed the observer a full range of responses and have witnessed the greater usefulness of the resulting ratio scales. It may be true that in the long run, superior procedures tend to replace inferior procedures, but in almost two decades of practice, magnitude estimation has not displaced category estimation—nor does it seem likely to do so any time soon.

Even among those who have given up category scaling in favor of ratio or magnitude matching, there remains a matter of taste and opinion that divides the practitioners. One view holds that the construct we call sensory magnitude follows simple laws, and in particular that under proper circumstances, the sensation magnitude experienced by the typical (median) observer grows as a power function of the stimulus magnitude. The power law is thought to set constraints on our expectations regarding the outcome of experiments, so that a discordant result becomes suspect until verified by adequate replication. In other words, the sensory power law takes precedence over the results of any particular experiment, just as the power law governing gravitational attraction usually remains unquestioned despite a particular experimenter's inability to confirm it by dropping objects in a laboratory. Both kinds of power laws make possible many kinds of predictions; yet they both can be shown to fail to some degree in particular contextual circumstances.

The other view holds that the nomothetic imperative has no compelling jurisdiction in psychophysics, because departures from the power law are too numerous to be ignored. It is claimed that the ways in which the human observer responds to stimulus intensity depend on prior learning, adaptation, range of stimuli, nature of the matching task, and so on; and that until those many factors and contextual influences can be discovered, explored, and understood, it is premature to speak of a psychophysical law.

The two views sketched above may prove more extreme than the attitude of any particular scientist, but they illustrate two poles of opinion. Perhaps the least nomothetic view yet expressed was that of Poulton (1968) who concluded his review of "the new psychophysics" by saying, "The mechanism of response learning and response bias must be included in any adequate description. To this reviewer," he added, "they present the more interesting and challenging problems." In my own view, the problems of response bias rate no better than a nuisance, an interesting nuisance, perhaps, as some of the foregoing sections have tried to show, but nevertheless a diversion from the basic business of sorting out the fundamental principles. Fortunately for science, however, its practitioners are motivated by a diversity of values and interests. It would stifle the enterprise if we all tried to crowd in on one single problem.

REFERENCES

BAIRD, J. C. *Psychophysical analysis of visual space.* Oxford: Pergamon Press, 1970.

Beck, J., & Shaw, W. A. Ratio-estimations of loudness-intervals. *American Journal of Psychology,* 1967, **80**, 59–65.

Berglund, B., Berglund, U., Engen, T., & Ekman, G. Individual psychophysical functions for twenty-eight odorants. *Perception and Psychophysics,* 1971, **9**, 379–384.

Dawson, W. E. An experimental analysis of judgments of sensory difference. Unpublished doctoral dissertation, Harvard University, 1968.

Ekman, G. Two generalized ratio scaling methods. *Journal of Psychology,* 1958, **45**, 287–295.

Fagot, R. A. On the psychophysical law and estimation procedures in psychophysical scaling. *Psychometrika,* 1963, **28**, 145–160.

Garner, W. R. A technique and a scale for loudness measurement *Journal of the Acoustical Society of America,* 1954, **26**, 73–88.

Geiger, P. H., & Firestone, F. A. The estimation of fractional loudness. *Journal of the Acoustical Society of America,* 1933, **5**, 25–30.

Ham, L. B., & Parkinson, J. S. Loudness and intensity relations. *Journal of the Acoustical Society of America,* 1932, **3**, 511–534.

Hellman, R. P., & Zwislocki, J. J. Loudness determinations at low sound frequencies. *Journal of the Acoustical Society of America,* 1968, **43**, 60–64.

Helson, H. *Adaptation-level theory: An experimental and systematic approach to behavior.* New York: Harper & Row, 1964.

Indow, T., & Stevens, S. S. Scaling of saturation and hue. *Perception and Psychophysics,* 1966, **1**, 253–272.

Jones, F. N., & Marcus, M. J. The subject effects in judgments of subjective magnitude. *Journal of Experimental Psychology,* 1961, **61**, 40–44.

Krantz, D. H. *A theory of magnitude estimation and cross-modality matching.* (MMPP 70-6) Ann Arbor: Michigan Mathematical Psychology Program, July 1970.

Krueger, L. E. Apparent combined length of two-line and four-line sets. *Perception and Psychophysics,* 1970, **8**, 210–214.

Marks, L. E. Stimulus-range, number of categories, and form of the category-scale. *American Journal of Psychology,* 1968, **81**, 467–479.

Merkel, J. Die Abhängigkeit zwischen Reiz und Empfindung. *Philosophische Studiën,* 1888, **4**, 541–594.

Miller, G. A. Sensitivity to changes in the intensity of white noise and its relation to masking and loudness. *Journal of the Acoustical Society of America,* 1947, **19**, 609–619.

Moskowitz, H. R. Scales of intensity for single and compound tastes. Unpublished doctoral dissertation, Harvard University, 1969.

Newhall, S. M. A method of evaluating the spacing of visual scales. *American Journal of Psychology,* 1950, **63**, 221–228.

Plateau, J. A. F. Sur le mesure des sensations physiques, et sur la loi qui lie l'intensité de ces sensations à l'intensité de la cause excitante. *Bulletin de l'Academie Royale de Belgique,* 1872, **33**(Ser. 2), 376–388.

Pollack, I. Iterative techniques for unbiased rating scales. *Quarterly Journal of Experimental Psychology,* 1965, **17**, 139–148.

Poulton, E. C. The new psychophysics: Six models for magnitude estimation. *Psychological Bulletin,* 1968, **69**, 1–19.

Poulton, E. C. Choice of first variable for single and repeated multiple estimates of loudness. *Journal of Experimental Psychology,* 1969, **80**, 249–253.

Savage, C. W. *The measurement of sensation.* Berkeley: University of California Press, 1970.

Stevens, J. C., & Guirao, M. Individual loudness functions. *Journal of the Acoustical Society of America,* 1964, **36**, 2210–2213.

Stevens, J. C., Mack, J. D., & Stevens, S. S. Growth of sensation on seven continua as measured by force of handgrip. *Journal of Experimental Psychology,* 1960, **59**, 60–67.

Stevens, J. C., & Marks, L. E. Spatial summation and the dynamics of warmth sensation. *Perception and Psychophysics,* 1971, **9**, 391–398.

Stevens, J. C., & Stevens, S. S. Brightness function: Effects of adaptation. *Journal of the Optical Society of America,* 1963, **53**, 375–385.

Stevens, J. C., & Tulving, E. Estimations of loudness by a group of untrained observers. *American Journal of Psychology,* 1957, **70**, 600–605.

Stevens, S. S. On the theory of scales of measurement. *Science,* 1946, **103**, 677–680.

Stevens, S. S. On the brightness of lights and the loudness of sounds. *Science,* 1953, **118**, 576. (Abstract)

Stevens, S. S. Biological transducers. *Convention Record, Institute of Radio Engineers,* 1954, Part 9, 27–33.

Stevens, S. S. The measurement of loudness. *Journal of the Acoustical Society of America,* 1955, **27**, 815–820. (a)

Stevens, S. S. On the averaging of data. *Science,* 1955, **121**, 113–116. (b)

Stevens, S. S. The direct estimation of sensory magnitudes—loudness. *American Journal of Psychology,* 1956, **69**, 1–25.

Stevens, S. S. Concerning the form of the loudness function. *Journal of the Acoustical Society of America,* 1957, **29**, 603–606. (a)

Stevens, S. S. On the psychophysical law. *Psychological Review,* 1957, **64**, 153–181 (b)

Stevens, S. S. On the new psychophysics. *Scandinavian Journal of Psychology,* 1960, **1**, 27–35.

Stevens, S. S. The psychophysics of sensory function. In W. A. Rosenblith (Ed.), *Sensory communication.* Cambridge, Mass.: M.I.T. Press, 1961. (a)

Stevens, S. S. To honor Fechner and repeal his law. *Science,* 1961, **133**, 80–86. (b)

Stevens, S. S. Power-group transformations under glare, masking, and recruitment. *Journal of the Acoustical Society of America,* 1966, **39**, 725–735.

STEVENS, S. S. Tactile vibration: Change of exponent with frequency. *Perception and Psychophysics,* 1968, **3**, 223–228. (a)

STEVENS, S. S. Measurement, statistics, and the schemapiric view. *Science,* 1968, **161**, 849–856. (b)

STEVENS, S. S. On predicting exponents for cross-modality matches. *Perception and Psychophysics,* 1969, **6**, 251–256.

STEVENS, S. S. Neural events and the psychophysical law. *Science,* 1970, **170**, 1043–1050.

STEVENS, S. S. Sensory power functions and neural events. In W. R. Loewenstein (Ed.), *Handbook of sensory physiology.* Vol. 1. New York: Springer-Verlag, 1971.

STEVENS, S. S. Perceived level of noise by Mark VII and dB(E). *Journal of the Acoustical Society of America,* in press.

STEVENS, S. S., & GALANTER, E. H. Ratio scales and category scales for a dozen perceptual continua. *Journal of Experimental Psychology,* 1957 **54**, 377–411.

STEVENS, S. S., & GREENBAUM, H. B. Regression effect in psychophysical judgment. *Perception and Psychophysics,* 1966, **1**, 439–446.

STEVENS, S. S., & GUIRAO, M. Loudness, reciprocality, and partition scales. *Journal of the Acoustical Society of America,* 1962, **34**, 1466–1471.

STEVENS, S. S., & GUIRAO, M. Subjective scaling of length and area and the matching of length to loudness and brightness. *Journal of Experimental Psychology,* 1963, **66**, 177–186.

STEVENS, S. S., & GUIRAO, M. Loudness functions under inhibition. *Perception and Psychophysics,* 1967, **2**, 459–465.

STEVENS, S. S., & POULTON, E. C. The estimation of loudness by unpracticed observers. *Journal of Experimental Psychology,* 1956, **51**, 71–78.

STEVENS, S. S., & STEVENS, J. C. *The dynamics of visual brightness.* Cambridge: Harvard University, Laboratory of Psychophysics, 1960.

TEGHTSOONIAN, R. On the exponents in Stevens' law and the constant in Ekman's law. *Psychological Review,* 1971, **78**, 71–80.

TORGERSON, W. S. Distances and ratios in psychophysical scaling. *Acta Psychologica,* 1961, **19**, 201–205.

(Received January 8, 1971; revision received May 6, 1971)

Psychological Review
1971, Vol. 78, No. 5, 451–456

TOWARD A THEORY OF EARLY INFANTILE AUTISM

DEWEY J. MOORE[1] AND DAVID A. SHIEK
Indiana State University *Western Kentucky University*

This paper hypothesizes that in early infantile autism, fetuses with an inherited high potential for intellectual superiority are developmentally ready for imprinting with human behaviors much before birth. Instead of being imprinted postnatally at critical periods in cerebral development as are most children, these neurologically advanced fetuses are imprinted on the restricted uterine environment. Such atypical imprinting leads to the syndrome of early infantile autism. Suggestions for research are offered.

Since Kanner (1943) wrote his paper in which he described a group of children who were characterized by "extreme autistic aloneness," the construct of early infantile autism has become engulfed in a morass of semantic confusion. Although Kanner believed that these children were special cases included in a larger group of children with childhood schizophrenia, subsequent investigations have used the terms schizophrenia, psychosis, atypical child, brain injured, and many others, to describe Kanner's group. The semantic problem has become further confounded by the fact that experimenters have purported to be investigating autism when their subjects were not at all like those Kanner discussed. For example, a study may report that a particular drug is useful with autistics when in fact the investigation failed to include a single child in the study who met Kanner's criteria.

The remainder of this paper, then, will discuss children designated by Kanner as being characterized by (*a*) inability to relate to people from the *beginning of life,* (*b*) failure to use language to communicate, (*c*) an obsession with maintaining sameness, (*d*) preoccupation with objects, and (*e*) evidence of good cognitive potentialities.

PARENTAL INTELLIGENCE

Based largely on his observation of their educational and social status, Kanner (1965) inferred that the parents of autistics are often remarkably intelligent. His statements have been questioned on the basis that wealthier, better educated, and presumably more intelligent parents are more likely to seek psychiatry than those of lower status (Bettelheim, 1967; Creak & Ini, 1960).

In spite of the doubts regarding Kanner's beliefs concerning the intelligence of parents of autistic children, evidence is accumulating to support his contention. For example, Rimland's (1964) careful analysis of the American-European literature containing descriptions of the parents of autistic children led him to state that the evidence overwhelmingly supports Kanner's early findings. Rutter (1969) points out that there are significant differences between the social class status of autistics and schizophrenics. In an epidemiological study, Lotter (1967) found that a high proportion of parents of autistic children are above average in intelligence and of superior socioeconomic status. Lotter is careful to show that this finding is not an artifact of referral policies, and is supported by Treffert (1970) who has found that highly educated parents do not identify their children as disturbed more often than less-well-educated parents.

Recent data from Treffert (1970) further supports the inference of high parental intelligence based on educational achievement. In his longitudinal study of 280 cases of childhood schizophrenia and infantile autism, Treffert found the educational level of parents with autistic children to be significantly higher than that of parents with schizophrenic children. Treffert states that his verification of this persistent finding

> could well point toward a genetic view of the illness with the autistic child as seen as the mutant of especially intelligent progenitors [p. 434].

---

[1] Requests for reprints should be sent to Dewey J. Moore, Porter Evaluation Center, Indiana State University, Terre Haute, Indiana 47809.

Not all studies support the suggestion of high parental intelligence (Levine & Olson, 1968; Lowe, 1966; Schain & Yannet, 1960), but these, as well as similar investigations, are difficult to evaluate because of their failure to adhere to the definition of autism as proposed by Kanner. Included in some of these studies are chronic undifferentiated schizophrenics and autistic/symbiotics (Lowe, 1966), autistic "mental defectives" (Schain & Yannet, 1960), and undifferentiated psychotics (Creak & Ini, 1960).

When investigators are careful to differentiate autism from other conditions that may simulate it, and when referral policies to clinics are properly accounted for, evidence is emerging that the inference of parental intellectual superiority is justified. Treffert's (1970) hypothesis that autistic children are mutants of highly intelligent parents appears entirely plausible. Interestingly, such a hypothesis lends credence to Kanner's (1943) observation that autistic children show evidence of good cognitive potentialities, and Rimland's (1964) review of the literature in which the finding of very special memory and manipulative abilities in autistics was commonplace.

NEURAL DYSFUNCTION

There is substantial growing support for the suggestion that early infantile autism is associated with neurophysiological dysfunction of the reticular activating system. Working with autistic children who were selected on the basis of Kanner's criteria, Hutt and Hutt (1969) found that electroencephalogram (EEG) records of their subjects showed desynchronized activity. These findings, together with behavioral observations, led the investigators to hypothesize that the children were in a condition of high physiological and behavioral arousal. Ornitz and Ritvo (1969) postulate that the basic condition causing autism rests on faulty homeostatic regulation of perceptual input. Indirect evidence by Holzman (1969) shows that dysfunction in the final or proprioceptive-autonomic response portion of the total perceptual process may be a significant factor in the etiology of the schizophrenic syndrome. Finally, Mirsky (1969) found 13 recent studies supporting the assumption that all schizophrenia begins with a basic fault in the ascending reticular activating system.

BRAIN DAMAGE

While many investigators agree that autism has a neurophysiological basis, probably in the reticular activating system, and that a potentially superior organism is involved, the interaction of these two findings needs exploration.

Rimland (1964), who subscribes to both of the above concepts, neurophysiological basis and organismic superiority, states,

> If a potentially very intelligent child whose brain was continuing to develop at a rapid rate, were exposed to atmospheric oxygen (or worse yet, to relatively high concentrations of medical oxygen) while the vasculature was still immature, damage to the vasculature of his brain could result [p. 129].

Rimland's (1964) theory of oxygen insult crumbles under the simple logic that if the child is genetically superior and developing at a rapid rate, it, even more than a normally growing organism, should be able to withstand normal stresses. The circumstances of a superior, rapidly developing brain that is *more* sensitive to normal environmental insults seems contradictory and denies all we would logically hypothesize about the ability of superior biological organisms to survive physical trauma.

A number of investigations have suggested that autistic children are more likely to evidence various kinds of brain pathology than are other children. Even when it is indicated that the researchers have taken great care to differentiate autistics from other behaviorally impaired children, the findings are inconclusive. As Rutter (1969) points out,

> There are no published histopathological studies; the EEG studies are inadequate and contradictory; and metabolic investigations are still in their infancy [p. 397].

Experimenters who purport to find brain damage in autistic children may actually be observing dysfunction of the reticular activating system. As will be shown later in this paper, such atypical functioning is not necessarily a result of brain damage but can result from sensory deprivation.

A review of "brain damage" as an etiologic factor in the production of autism is of little value. Rutter (1969) states,

> However, even if it is accepted that autism may sometimes, or perhaps often, develop on the basis of brain damage, the explanation is too general to be of much help in understanding how autism develops. The term "brain damage" includes a wide variety of disturbances. The vast majority of children with brain damage are not autistic, and it remains to be explained why the minority who become autistic do so. More specific biological hypotheses are required [p. 397].

SENSORY RESTRICTION

Until very recently, psychologists have largely ignored the prenatal *psychological* environment. The assumption seemed to have been in the past that all children start from a zero point at the moment of birth and few significant environmental influences, other than those affecting physical growth and development, have made an impression up until that time. Such an assumption has remarkable commonsense validity in view of the limited sensory experiences available to the fetus, but the question remains as to whether this restricted environment is the same for all fetuses. Does the prenatal psychological environment have a different effect on the cerebral structures of superior, rapidly developing fetuses than it does on those who develop more normally?

Part of the answer to this question is supplied by experiments concerned with the effects of sensory restriction. Schultz (1965) comments,

> restriction of early environmental stimulation in animals has produced drastic and enduring effects on emotionality, learning ability, activity level, social behavior, and perception. The optimal level of sensory variation of the individual is influenced by early postnatal levels of stimulation and the resulting level of cortical arousal as mediated by the reticular activating system. In other words, the level of stimulus variation to which the neonate is exposed functions to influence the optimal range of cortical arousal appropriate for adaptive behavior of the adult organism. The adaptation level of the ascending reticular activating system . . . is thus determined to some degree by the amount of early stimulation [pp. 25–26].

Schultz suggests that an impoverished sensory environment would prevent differentiation of cue functions, sensory modalities, and of events within modalities. Such a failure to differentiate leads to defects in the perceptual process.

Evidence that autism may result from sensory deprivation is supported by Rimland's (1964) review of the literature in which he concludes,

> there is obviously a good deal of behavioral similarity between the victims of early infantile autism and the victims of sensory deprivation [p. 104].

He notes that subjects who experience long periods of sensory deprivation are "detached from the external environment," and that they "stare right at you but never see you and just sit and look into space and not talk [p. 104]." The vacant "long eye" stare usually observed in these sensorially deprived subjects closely resembles the unfocused gaze of autistic children. Furthermore, Rimland's (1964) search of the literature has shown that the reticular formation is the most likely site in the brain to be associated with the phenomenon of sensory deprivation. He explains that there is sufficient reason for believing that autistic children are so perceptually impaired that sensory deprivation psychosis is entirely within the realm of possibility.

DEVELOPMENTAL READINESS

An additional concept needs to be considered in order to answer the question: Does prenatal psychological environment have a different effect on the cerebral structures of superior, rapidly developing fetuses than it does on those who develop more normally? This additional concept has to do with the developmental readiness of the fetal brain.

Differences in the maturational levels of full-term neonates as evidenced by variation in physical size, activity level, and the presence of developmental landmarks have been noted for decades. These same maturational differences exist prior to birth. Some fetuses have cerebral structures that are developmentally ready for sensory stimulation long before others. The neurophysiology of a fetus that is the product of intellectually superior parents can reasonably be expected to show advanced maturation. As Ames (1969) has suggested, intellectual superiority tends to go along with advanced maturity.

The quantity and quality of early stimulation have enduring effects on the functioning of the brain. Profound environmental deprivation produces organic dysfunction (Rutter, 1969) and may *simulate* "brain damage." Significantly, the negative effects of such environmental deprivation are relatively greater for genetically brighter organisms (Schultz, 1965).

Given a fetus with a brain in an advanced state of developmental readiness for stimulation but residing within a restricted uterine environment, what would be the probable results? What are the effects of sensory restriction and failure to experience primary socialization when these two circumstances occur simultaneously?

PRIMARY SOCIALIZATION AND IMPRINTING

The striking similarities between the behavior of autistic children and that of nonimprinted or artificially imprinted animals are too great to ignore. The autistic aloneness of the child and his lack of interest in his own species has many parallels in studies demonstrating the species-ignoring behavior of animals that are not imprinted with their own kind. Writing about his experiments with geese, Lorenz (1957) says,

> Once their social relations are transposed to a human being their behavior does not change even if they are kept for years with other members of their species . . . [p. 105].

Moltz (1960) points out that once imprinting occurs, its effect cannot be reversed. The stimulus the animal experiences during the critical imprinting period becomes either the preferred stimulus or the only stimulus toward which the animal will move. Scott (1969) states in his discussion of the primary socialization in canine and human infants that an individual at the proper period in life will become attached to anything in the surrounding environment, both living and nonliving.

Addressing himself to the importance of the critical period concepts, Money (1969) writes,

> The essential elements of the concept are: 1) that a developmental readiness of the organism must be met by a suitable stimulus environment before the given phase of development can take place; 2) there is a time limit, after which the developmental opportunity is lost, often forever; 3) the events of the developmental phase leave a permanent, or at least a very durable, residue; 4) defective or impaired development may leave a permanent residual effect; and 5) as applied to social science, there is no clear distinction between body and mind for the events of the social life as assimilated by the brain which is affected by what it assimilates as well as effecting in turn, the events outside of itself in social life [p. 145].

Does the autistic child's insistence on sameness in the environment, his repetitive and stereotyped behaviors, his hand regarding, and his failure to develop speech suggest that he was imprinted by an atypical environment at a critical period in his primary socialization? In this essentially unchanging environment, the fetus has only the heartbeats of himself and his mother, and perhaps the other rhythms of their bodies, to affect him. Is it possible that the autistic child is so firmly imprinted with his relatively featureless, secluded, and limited world during critical developmental periods, when he should be vividly stimulated and experiencing primary socialization, that he in fact seems to be living "in his own world" after birth? Does it explain his apparent need to maintain environmental sameness?

What are the effects of rhythmic heartbeats and body movements on a developmentally superior fetus? Do these rhythms account for the repetitive and stereotyped behaviors observed later in life? Does such "rhythmic imprinting" explain Rimland's (1964) findings that astonishing musical ability has been reported in at least nine studies of autistic children?

Is the fetus "kinesthetically imprinted" on his own hands? They constitute the major movable stimuli in his dark prenatal world. Does this account for the frequently observed exceptional manipulative skills shown after birth?

Rutter (1969) notes that at 16 weeks, the normal infant is definitely responding meaningfully to human speech. What happens when a neurologically advanced, developmentally ready fetus does not receive early speech stimulation because it is unavailable to him?

Psychological literature contains many reports of feral children who remained wild after many years of association with humans. Was their inability to assume human characteristics due to a failure to experience primary socialization with humans at a critical period in their development?

Spitz (1965), reporting on the work of M. Von Senden, notes that several congenitally blind persons whose sight was restored between the ages of 3 to 40 never learned to "see" although they had vision. Had the critical period for imprinting for meaningful visual perceptions passed? Was their behavior similar to autistic children who "look through" persons but never "at" them? Do humans *as* humans have any meaning for autistic children, or are they similar to lower animals who have been prevented from being imprinted on their own species?

Are autistic children, by their frequent obsessions with inanimate objects, demonstrating Scott's (1969) finding that an individual at the proper time in life will become attached to both living and nonliving things in the surrounding environment?

The period for primary socialization in the human infant extends from approximately six weeks to six months after birth. It is during this time that the infant forms his first meaningful responses to his environment (Scott, 1969). This is the critical period for much of his important species imprinting. What happens when a fetus, who has inherited superior neurological potential from his parents, approaches his ninth month? He is hardly an "average" fetus. If his eventual IQ is projected to be 170 or more, his unstimulated brain has already "passed through" the first six months of what would ordinarily be his postnatal life. Consequently the optimal period for his primary socialization has passed.

Are autistic children "mutants of especially intelligent progenitors" as Treffert (1970) suggests they may be? Are autistics not just superior organisms, but do they have potential intellectual endowment so rare that it might occur in only the upper extreme limits of the distribution of intelligence? Interestingly, the prevalence of infantile autism is approximately 2 per 10,000 (Treffert, 1970). This prevalence is almost identical to the number of persons thought to have rates of intellectual development which might be expressed as quotients of more than 170 if tests were available with which we could measure them.

SUGGESTIONS FOR RESEARCH

The most obvious research need for testing the hypotheses stated in this paper revolves around the necessity for determining the intellectual levels of parents who have autistic children. These investigations should follow Kanner's definition of autism carefully to avoid contaminating the results with scores of parents whose children have "autistic symptoms" but are not truly early infantile autistics. It may be necessary for researchers to pay particular attention to the quality of parental intellectual activity as well as the quantity. This seems especially advisable in view of Kanner's (1965) observation that the parents of autistic children, while they may have meager incomes, are preoccupied with abstractions of a scientific, literary, artistic, or mathematical nature.

More research is needed to test the neurophysiological aspects of autism. In this respect, the work by Hutt and Hutt (1969) mentioned earlier seems an especially appropriate area of investigation. Their experiments led them to hypothesize that autistic children are in a high state of physiological and behavioral arousal. This paper has suggested that in autism it is believed that the fetus becomes habituated to a low level of stimulus variations. Reference was made previously to Schultz's (1965) statement that the level of stimulus variation the organism experiences early in its development is an important determinant of the optimal cortical arousal for adaptive behavior. Because there is little stimulation prenatally, the child's optimal cortical arousal is low. After birth, normal levels of environmental stimulation constitute an unaccustomed overload with consequent cortical hyperarousal.

1968; Shepard, 1964; Thurstone & Jones, 1957) has been based on the assumption that the numerical values of the dependent variable can be regarded as interval-scale measures of the relevant psychological attribute." The subject of the sentence is not Anderson's work; his paper was mentioned, along with other reviews, because it cites many studies in which the interval-scale assumption was utilized.

Anderson has indeed discussed the possibility of monotone transformations of the dependent variable. However, his treatment of this question is inadequate: He presents no general methods for testing whether appropriate transformations exist, and consequently can reach wrong conclusions (e.g., Sidowski & Anderson, 1967).

We agree that numerical responses often do contain useful "metric" information. When and why they do so is itself a subject for qualitative investigation. Stevens' methods of magnitude estimation and cross-modality matching, for example, lend themselves to analysis using conjoint-measurement methods.

One final remark: Anderson describes his work by inserting adjectives and adverbs with positive affective loadings but little precise content, for example, "organically" related and "functional" measurement. He declares it to be "superior on epistemological grounds," "an organic part of behavior theory," and "a clear rationale and a basis for the general study of judgment." Such assertions are very poor substitutes for adequate reasoning about the foundations of psychological measurement.

REFERENCES

Krantz, D. H., & Tversky, A. Conjoint-measurement analysis of composition rules in psychology. *Psychological Review,* 1971, **78**, 151–169.

Sidowski, J. B., & Anderson, N. H. Judgments of city-occupation combinations. *Psychonomic Science*, 1967, **7**, 279–280.

(Received May 28, 1971)

David H. Krantz
Department of Psychology
University of Michigan
Ann Arbor, Michigan 48104
and
Amos Tversky
Hebrew University
Jerusalem, Israel

Krantz and Tversky apparently agree with the main point of my letter, that functional measurement does not assume an equal-interval response scale. Now, however, they claim that my treatment of response scaling is "inadequate," which I should say is incorrect. The power series method (e.g., Anderson, 1970, p. 159) has limitations, of course, but it has some definite advantages over nonmetric methods, especially for "fallible," that is, real data.

Krantz and Tversky are seriously in error in stating that I ignore qualitative laws. As one example, qualitative analysis has provided critical support for the averaging hypothesis in repeated experiments (see Anderson, 1971, p. 181). This finding has important consequences for measurement theory.

Krantz and Tversky also claim that I quoted them out of context. The cogency of the argument in their letter (that their reference to Anderson, 1970, should not have been taken as a reference to my work) is perhaps better judged by others. For my own part, I wrote them two or three times concerning this same statement in their earlier technical report; they could easily have straightened out the matter had they chosen to do so.

REFERENCES

Anderson, N. H. Functional measurement and psychophysical judgment. *Psychological Review*, 1970, **77**, 153–170.

Anderson, N. H. Integration theory and attitude change. *Psychological Review*, 1971, **78**, 171–206.

(Received June 11, 1971)

Norman H. Anderson

Our remarks about the advantages of qualitative analysis have been tacitly accepted by Anderson. The systematic study of qualitative laws is a major part of the subject matter of conjoint measurement. We hope, therefore, that conjoint-measurement methods will henceforth be used more systematically by Anderson in the analysis of composition rules and also of response scales.

(Received June 17, 1971)

David H. Krantz
and
Amos Tversky

29 MAR 1972

to-

15/11/71

PSYCHOLOGICAL REVIEW

GEORGE MANDLER, EDITOR
JEAN M. MANDLER, ASSOCIATE EDITOR
UNIVERSITY OF CALIFORNIA, SAN DIEGO

Vol. 78, No. 4 July 1971

Articles

On Perceptual Clarity.................P. C. Dodwell 275

Organization and Interference..........Leo Postman 290

Function of Distinct Associations for Paired-Associate Performance......................George Wolford 303

Verbal Learning Theory and Independent Retrieval Phenomena......................Edwin Martin 314

Theoretical Notes

Self-Punitive Behavior—Masochism or Confusion?
Paul Dreyer and K. Edward Renner 333

Strength Theory and Confidence Ratings in Recall
Harley A. Bernbach 338

The Behavioristic Conception of Neurosis: A Reply to Two Critics.......................Joseph Wolpe 341

Avoidance Conditioning and Behavior Therapies: A Reply to Costello
Arnold Powell and Augustus R. Lumia 344

Conditioned Avoidance Responses and Phobias: A Reply to Wolpe and to Powell and Lumia
C. G. Costello 348

A Reply to Wolpe...............Bernard Weitzman 352

Letters to the Editor 354

Editorial Note......................... 289

PSYCHOLOGICAL REVIEW

Published Bimonthly
by The American Psychological Association, Inc.

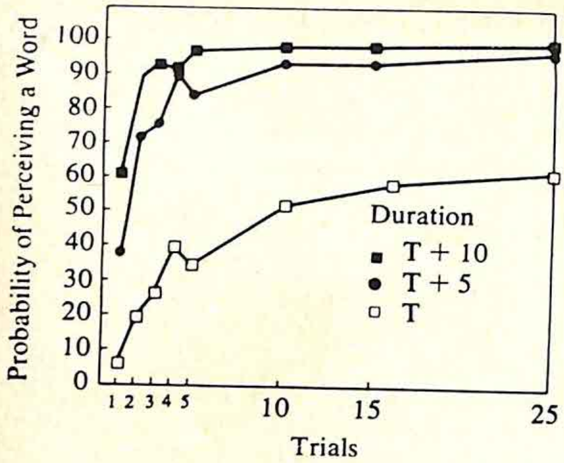


FIG. 1. Probability of perceiving a word as a function of repetition trials (1 through 25) and duration. (The base duration, T, was found individually for each subject and was typically about 15 milliseconds.) (From Haber, 1969; originally in Haber and Hershenson, 1965. Reprinted from an article by R. N. Haber and M. Hershenson published in the January 1965 *Journal of Experimental Psychology.* Copyrighted by the American Psychological Association, Inc.)

principle is not broad enough to account for integration of information over changes which are not a direct result of self-produced movement of the receptors. In what follows, a model is proposed which goes some way toward suggesting a basis for visual storage and sequential processing in the sense of integrating stored visual information with new inputs, and hence a basis for perceptual stability and clarity.

REPETITION AND PERCEPTION

Perhaps the most intriguing recent finding on visual storage is Haber's clarity effect: an observer who views a briefly presented stimulus array repeatedly, and at constant luminance and duration for each repetition, reports that successive presentations increase in perceptual clarity. Obviously, the effect is obtainable only over a restricted range of conditions; the presentations must not be strong enough to yield complete clarity on the first exposure, nor so weak as to yield no perception at all after repeated exposures. Nevertheless, it is remarkable that repetition per se can have such a powerful effect. The finding has been well documented under a variety of conditions, recently summarized by Haber (1969).

The major results are shown in Figures 1 and 2. The experiments have all been concerned with seven-letter arrays that spell words. The first figure shows how exposure duration affects clarity, the second that both familiarity and prior knowledge of what the word is increases the probability of its being seen, but not the change in clarity from trial to trial. It is very unlikely that guessing is an important factor in the experiments, since prior knowledge of the word does not eliminate the repetition effect. Taking the two figures together, it appears that the various experimental manipulations (exposure duration, familiarity, prior knowledge) determine both the initial and final levels of performance, but *not* the rate at which clarity grows with repetition, a point Haber himself makes with some force (Haber, 1969, p. 277).

How may such a finding be explained? Haber favors an explanation in terms of the arousal of Hebbian cell assemblies. In a general way, this seems reasonable, since Hebb is one of the few theorists who has considered the role of repetition in percep-

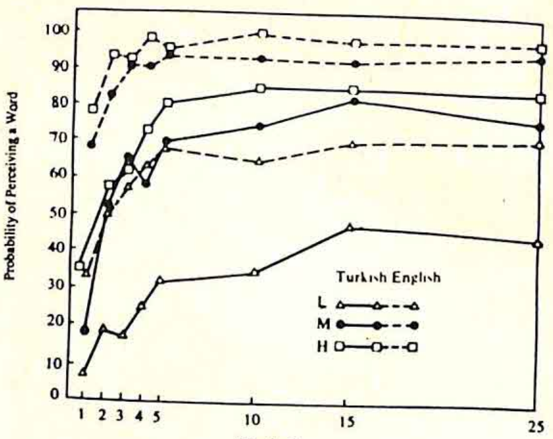


FIG. 2. Probability of perceiving a word as a function of repetition trials, duration (low—15 milliseconds; middle—20 milliseconds; high—25 milliseconds), and language (English and Turkish). (From Haber, 1969; originally in Hershenson and Haber, 1965. Reprinted with permission from an article by M. Hershenson and R. N. Haber published in the *Canadian Journal of Psychology,* 1965, Vol. 19. Copyrighted by the Canadian Psychological Association, 1965.)

tual organization in any systematic way. The problem with the cell assembly, however, is that it is too vague and general a concept; it will oblige by performing almost any job assigned to it. In this particular case, it can be used both to explain the growth in clarity (successive arousals of an assembly lead to its firing more strongly) and also the different levels reached for different sorts of words (better organized assemblies fire more strongly than less-well-organized ones). However, one should then expect rate of growth and final clarity level to be correlated, since both would be functions of prior strength of the relevant cell assembly. The data, per contra, suggest something different. It seems that level of clarity is affected by differences in familiarity and in signal strength or duration, but rate of growth of clarity is not. The differences between the sets of curves in Figures 1 and 2 can be accounted for by variation in a single parameter which specifies asymptotic performance. I propose that a second, constant, growth parameter has to do with signal processing at a lower level of information extraction than that at which word familiarity has its effect. Notice that the units of all the displays are similar—printed letters—so that a constant parameter which has to do with initial processing of the elements for spatial–structural information would hardly be surprising. The hypothesis is that the signal processing involved is a form of autocorrelation which determines the growth rate parameter.

This hypothesis may seem to be a gratuitous leap beyond the evidence; I shall justify it by showing that there are several other features of detail vision and perceptual clarity which support the notion, which indeed lend it quite strong confirmation. Two have to do with major puzzles in pattern vision: the first of these is the disappearance of retinally stabilized images, and their regeneration in nonrandom ways; the acuity of the eyes (particularly binocular acuity) in the presence of relatively large nystagmic movements and fixation errors is the other. Neither phenomenon has heretofore been explained adequately. Third, I shall show that neurophysiological findings on contour processing in the mammalian visual system suggest a very plausible basis for the limited form of autocorrelation function required by the model. Fourth, I shall show how an autocorrelation model solves rather elegantly the problem of the detection of signals in noise, given the physical and physiological constraints under which the visual system normally operates, and the evolutionary pressures to which it has been subjected. The biological utility of the proposed form of signal processing will thus be demonstrated. Finally, a correlational model will be shown to be a possible basis for the forms of information storage, integration, and interference discussed in the previous section.

CONVOLUTION THEORY AND PATTERN SIMILARITY

Several proposals have been made for employing autocorrelation functions, which are a particular type of convolution, in pattern recognition. MacLachlan (1962) suggested useful functions for them in optical and mechanical devices for information retrieval and pattern recognition, and they play a major role in recent work on picture processing by computer (see Rosenfeld, 1969). Engel (1969) has very successfully employed them in modeling the phenomena of binocular brightness mixing; indeed, it was his work which suggested to me the more general applications described in this paper.

The mathematics of convolution theory required for the present purpose are straightforward. Given a pair of continuous two-dimensional patterns $\rho(x,y)$ and $\rho'(x,y)$, their cross-correlation is defined by the functional

$$A(uv) = \int\int \rho(x,y)\,\rho'(x+u,\,y+v)\,dx\,dy, \quad [1a]$$

where u and v are new parameters which can be thought of as internal coordinates originating on one of the patterns. The complete expression of the cross-correlation is then:

$$\int\int A(uv)\,du\,dv \qquad [2a]$$

over the appropriate values of u and v.

For a discrete, or digitalized, pattern the analogous expressions are:

$$A(uv) = \sum_{x}\sum_{y} \rho(x,y)\ \rho'(x + u, y + v) \quad [1b]$$

and

$$\sum_{u}\sum_{v} A(uv). \quad [2b]$$

In either case, Expressions 1a and 1b state that the brightness values at all points (x, y) are multiplied by the brightness values at $(x + u, y + v)$ and the cross-products summed. Expressions 2a and 2b state that these quantities are summed for all values of u and v. In practice, some limitation of the range of values for u and v is likely, so that the operations become local, position invariant transformations of the original pattern (Rosenfeld, 1969, Ch. 4).

The autocorrelation of a pattern with itself is simply obtained by substituting $\rho(xy)$ for $\rho'(xy)$; that is, it is obtained from

$$A(uv) = \int\!\!\int \rho(xy)\rho(x + u, y + v)dxdy \quad [3a]$$

over the appropriate values of u and v. The corresponding expressions for discrete patterns are obvious.

For two-dimensional spatial patterns, Expressions 1a, 1b, 2a, and 2b have a simple interpretation, which is demonstrated in Figure 3. For given the values of u and v, $A(uv)$ is a measure of the overlap of two patterns; in the case of autocorrelation, it is the degree of overlap of a pattern with itself, at a given shift u along the horizontal (x) axis, and v along the vertical (y) axis. Integrating (or summing, in the case of discrete representation) the values of $A(uv)$ over the appropriate ranges of u and v will then give the measure $\int\!\!\int A(uv)dudv$ which expresses overall the cross-correlation or autocorrelation as the case may be. Clearly the autocorrelation of a pattern with itself has some characteristics not shared by a cross-correlation. In particular, there is an exceptionally high value of the function when $u = v = 0$, that is, where the pattern completely overlaps itself. This is illustrated in Figure 4, adapted from MacLachlan (1962). MacLachlan shows that $\int\!\!\int A(uv)dudv$ is not an appropriate measure of pattern similarity under all circumstances, but that $\int\!\!\int A^2(uv)dudv$ is (the discrete analogue being $\sum\sum A^2(uv)$). The second expression he calls the similarity function, which may be written $S(\rho\rho')$ and is the same as Engel's variance function $V(\rho\rho)$ in the case where the patterns are identical (i.e., when $A(uv)$ is an autocorrelation). Engel points out that $V(\rho\rho)$ is a good measure of the spatial–structural information in a pattern. $S(\rho\rho')$ similarly is a measure of shared structure between two patterns.

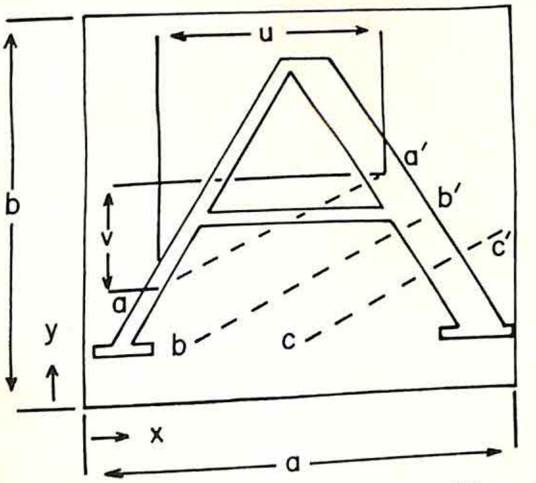


FIG. 3. The letter A in the domain ab. The vectors aa′, bb′, and cc′ all have the same length and orientation, so that u and v are the same for all. (The products of the densities, or brightnesses, at the termini, are the quantities of interest.) (Adapted from MacLachlan, 1962.)

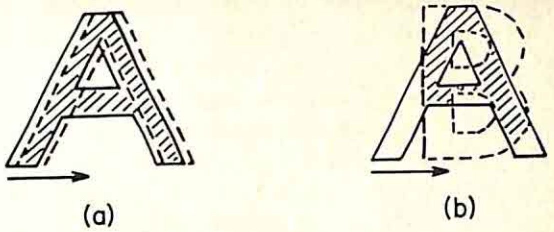


FIG. 4. When identical patterns, as in (a) are brought into coincidence $(u = v = o)$, the common area (shown hatched) is greater than that obtained with unlike patterns at maximum coincidence, as in (b).

So much for the formalism of convolutions: further information on autocorrelation and its application to picture processing may be found in Rosenfeld (1969, esp. Ch. 4, 5, and 6). If the ideas are to be applied to pattern processing in the visual system, it will be necessary to consider a few practical constraints which are suggested by the neurophysiology of the system, and also to propose a mechanism for the generation of the autocorrelation functions. This topic is taken up in detail in a later section; for the moment it may be noted that in his computer simulations of binocular brightness mixing, Engel was able to obtain remarkably good agreement with the human experimental data by using a relatively coarse approximation to $A(uv)$ and $V(\rho\rho)$. Perhaps, then, an indication of how fairly crude versions of these functions might be computed in the visual system will also be sufficient for our purposes.

Autocorrelations of two-dimensional spatial patterns are obtained by shifting a pattern with respect to itself and evaluating the amount of overlap for different degrees of shift. The similarity function $S(\rho\rho')$ is computed from the same basic data, and most of the useful information can be obtained by relatively small shifts, provided that contour information is extracted *before* the autocorrelation is performed. We know from the work of Hubel and Wiesel (1962, 1968) and others that contour information is abstracted at least at the first cortical processing level in mammals, and also that small shifts of the retinal pattern of stimulation with respect to itself are ensured under normal circumstances by elaborate patterns of rapid and involuntary eye movements (Krauskopf, Cornsweet, & Riggs, 1960; Ratliff & Riggs, 1950). Prima facie, then, there is evidence that mechanisms for computing some form of autocorrelation are present at least in higher visual systems.

IMAGE CLARITY AND AUTOCORRELATION

Repetition

To return to Haber's repetition effect, the specific hypothesis is that a necessary condition for the clear perception of a visual image is that the visual system perform an autocorrelation on the input array, a process which takes a finite amount of time, and that the similarity function $S(\rho\rho)$—or some close analogue of it—is also computed. The perceived clarity of the image is determined by $S(\rho\rho)$; it is at least a function that increases monotonically with $S(\rho\rho)$. The reason why brief tachistoscopic presentations of patterns do not lead to the perception of clear images is that insufficient time is available for the complete computation of $S(\rho\rho)$. If, however, values of $S(\rho\rho)$ (including incomplete ones) are stored for a short time, subsequent inputs of the *same* pattern can summate with the stored version, and in the latter case, can enter into its more complete specification.

Lack of clarity on first presentation of a letter array is, as Haber (1969) points out, the single best indicator of the final level of clarity that will be achieved under the given conditions. The autocorrelation model explains this by saying that lack of clarity on first presentation is an index of failure to complete the formation of $S(\rho\rho)$; the less $S(\rho\rho)$ can be specified on Trial 1, the less well can it *ever* be specified under the given stimulating conditions. If the model is on the right track, one would predict that the spatial–structural information in the image would affect the rate of growth of $S(\rho\rho)$ and hence of clarity, whereas signal strength (e.g., brightness) might be relatively ineffective. So far, Haber's experiments have all been with letter arrays, so that the type of spatial–structural information has been maintained relatively constant. Specific predictions about the effects to be expected in the repetition paradigm when stimulus information is varied systematically will be given later.

Image Stabilization and Disappearance

If the image of a target is stabilized on the retina, so that eye movements cause no relative shift between image and retina, the perceived target rapidly loses clarity and fades away (Ditchburn & Ginsborg, 1952; Riggs, Ratliff, Cornsweet, & Cornsweet, 1953). If the stabilization is complete, image fading is also complete and permanent (Barlow, 1963). This surprising phenomenon has never been adequately explained; Barlow proposed that there is a rapidly adapting information channel for detail vision which requires constant change of stimulus information to remain active, and several writers have remarked on the fact that the contour coding system based on the retinal receptive field organization of Hubel and Wiesel responds best to moving contours, and at higher levels perhaps *only* to moving contours. Heckenmueller (1965) points out that stabilization could lead to selective loss of function in one or more of the types of processing achieved at different levels of receptive field organization.

The specific hypothesis here is that image fading is fundamentally the obverse of the repetition clarity effect. In the latter, increases in clarity occur because successive inputs add increments of information leading to better specification of the autocorrelation function and $S(\rho\rho)$; in the former, there is successive *loss* of information on which the autocorrelation function can be computed. If the clear perception of a contoured pattern depends on the concurrent formation and maintenance of $S(\rho\rho)$, and if the generation and maintenance of $S(\rho\rho)$ depends on shifts of the pattern with respect to itself which are mediated by eye movements, the result follows obviously. The only problem then is to explain why the target is *ever* seen when the retinal image is stabilized, since by definition no shift of the pattern with respect to itself is possible. The problem is not insurmountable, however; a plausible extension of the notion of correlation on the input can handle it. This is presented in a later section.

Image Regeneration

The most complete information on image regeneration under stabilized—or near stabilized—conditions is contained in papers by Pritchard, Heron, and Hebb (1960) and by Evans (1965). The authors of both papers stress the nonrandom character of regeneration, and interpret their findings in terms of Gestalt-like processes in pattern perception and, in the former case, in terms of the cell-assembly model. In summarizing this work elsewhere (Dodwell, 1970, Ch. 9), it was pointed out that this nonrandomness of regeneration is in many respects similar to the nonrandomness found in rivalry between pattern elements during binocular stimulation of the eyes with noncongruent stereograms, and should be able to tell us something about the processing of patterns in the visual system. More definitely, one may postulate that at least some of the effects are caused by the way in which the autocorrelations and $S(\rho\rho)$ are computed. Both Pritchard et al. and Evans reported that "structured" disappearances are much more frequent than "random" or "unitary" ones; straight lines tend to fade and regenerate as units, pairs of parallel lines tend to fade together, and horizontal lines are more dominant than others. Patterns composed of curved contours tend to be less "active" than patterns consisting of collections of straight lines; the latter fade and regenerate more readily. These findings are consistent with the hypothesis that perceptual clarity is mediated by signal processing which is based on a form of autocorrelation, particularly if the computations are based on Hubel-and-Wiesel-type straight-line detectors. The detailed exposition of this hypothesis is presented later; here it should be noted that just as for the Haber clarity effect, a signal-processing model cannot explain all the findings because there are "higher level" effects which depend on attention and meaningfulness. Pritchard et al. report, for example, that meaningful targets (such as a face) tend to fade less readily than meaningless targets, and the fading seems to be determined to some degree by meaningful connections between parts of a target; an element fades less readily when the observer's attention is focused on it, and so on. Rather similar findings have also been reported by McKinney (1966). My argument here is the same as before; in criticizing the use of the cell-assembly model to explain the repetition effect, attention was drawn to the overinclusive nature of that model. Information processing is surely occurring at several levels, and the fading and regeneration effects in simple targets can be explained by signal processing at a level prior to that at which expectations, meaningfulness, and so on, have their impact.

Visual Acuity

It has been proposed (e.g., Andersen & Weymouth, 1923) that one function of "involuntary" eye movements is to increase acuity by stimulating large numbers of retinal points and thereby increase the amount of data on which a "mean retinal local sign" can be computed. Evidence on acuity for stabilized targets seems to contradict this notion. Riggs et al. (1953) showed that acuity for a stabilized target (before it fades) is actually somewhat better than in the nonstabilized condition. Further evidence reported by Keesey (1960) demonstrates that there are no consistent effects of eye movements on acuity, and that increase in acuity over time (typically between about .01 and 1.0 second) is similar for stabilized and nonstabilized targets. Shortess and Krauskopf (1961) report a similar finding for stereoacuity. Although eye movements are necessary for the maintenance of visual clarity, they do not themselves contribute to increasing the sharpness of detail vision.

Perhaps the most striking fact about the relations between acuity and eye movements is that reported by Krauskopf et al. (1960). Physiological nystagmus in the two eyes is uncorrelated; saccades tend to be yoked, but are not direct vergence adjusters. This means that even an experienced observer attempting to maintain steady binocular fixation is not able to do so with great accuracy. Krauskopf et al. measured the standard deviation of convergence error for two experienced observers at greater than two minutes of arc for sample durations of 60 seconds, and well over one minute of arc for one observer over a period of only 2 seconds. Yet, stereoacuity has been reported to be of the order of five *seconds* of arc. It seems clear that in order to extract information efficiently from the "noisy" analogues of the visual field provided by eyes subjected to constant tremor, drift, and saccades (Ratliff & Riggs, 1950), some form of statistical averaging, or smoothing, must occur. The present hypothesis is that a sufficient operation for this extraction can be based on the autocorrelation of a visual image with itself, and in the case of binocular vision, the cross-correlation of the images in the two eyes (Engel, 1969).

Specifically, the hypothesis accounts for the fact that both monocular and binocular acuity are very much finer than the vagaries of eye movements would lead one to expect. If the eye movements are thought of simply as a necessary condition for the computation of the autocorrelation and $S(\rho\rho)$, these functions need not themselves be related directly to the extent of the eye movements, and more probably depend principally on the grain of the retinal mosaic and the characteristics of the contour-coding units in the primary visual areas of the brain cortex.

Monocular acuity is initially at least as good for a stabilized as for an unstabilized target: let us postulate for the moment that there is a correlational process (detailed in the next section) equivalent to the autocorrelation on the initial input to account for that fact. It was argued earlier in connection with repetition and clarity that the signal processing must take a finite amount of time, and that maximum clarity (acuity) will only be achieved when the processing is complete. Although Haber claims that signal duration is not itself an important factor in the growth of perceived clarity, there is a suggestion in his data that duration may be traded off against repetition. This is illustrated in Figures 1 and 2; it could be argued that under his conditions, a total exposure of between 100 and 150 milliseconds is needed to attain full clarity. It would hardly be surprising if processing for fine details (e.g., in acuity determination) were to take several times as long as this, as in Keesey's experiment. However, this is straining the evidence; we can at least say that there is nothing grossly inconsistent about the time course of repetition clarity, of increases in acuity with time, and of stopped image fading.

A NEUROPHYSIOLOGICAL BASIS FOR AUTOCORRELATION

In keeping with the general idea that the basis for perceptual clarity effects should be sought "early on" in the visual processing system, it is next demonstrated that the known neurophysiology of contour coding provides a very plausible basis for the computation of autocorrelations. It has already been suggested that populations of Hubel-and-Wiesel-type detector units for edges and straight-line segments may fulfill the requisite conditions. There appears to be a surprising degree of redundancy in the coding arrangements, not all of which can be accounted for by the fact that some cortical units are binocular disparity detectors (Barlow, Blakemore, & Pettigrew, 1967; Pettigrew, Nikara, & Bishop, 1968). Cells are arranged in columns normal to the surface of the cortex in Areas 17, 18, and 19 in both cats and monkeys, nearly all cells within a column being selective for a contour in a specific orientation, but also within a fairly circumscribed region of the retina (Hubel & Wiesel, 1962, 1968). Most of the cells respond particularly well when the line segment to whose orientation they are tuned moves in a direction perpendicular to its orientation. Consider the somewhat idealized situation in which the external pattern and the head are both stationary, but the eyes move; consider also one pattern element, a straight line, which is in the right position and orientation to stimulate the cells of a particular column. That component of the eye movement which is orthogonal to the preferred orientation for the given column will in effect shift the line with respect to itself and cause brisk outputs from a subset of units within the column. As the amplitude of nystagmus varies, or as the eyes drift, the subset of strongly stimulated units will also change. But, given that within a column there are units with different types of receptive field ("on" center, "off" center, simple and complex, of different sizes, etc.), the *pattern* of firing would specify a stimulus element more uniquely than would any one unit and its associated receptive field alone. Here, then, we have a mechanism for signal processing of the proposed type. Stimulus detection is based on the statistical evaluation of the outputs from a population of units over space and time, an evaluation which can transcend the acuity or reliability of any one unit on its own. The process is one of shifting an element with respect to itself and continuously evaluating the result; one might say that, initially at least, this is a cross-correlation between the pattern element (line segment) and the set of receptive field units within a column. The output from this column constitutes "matched filtering" for the pattern element in question (Rosenfeld, 1969, pp. 71–78). The problem for the visual system now is to decide whether this filtered output arises from a stimulus pattern that is constant in time, or whether the input is changing. This is where autocorrelation becomes important. It is reasonable to talk of the autocorrelation of an image with itself where the input pattern is known to remain the same, but it seems probable that the visual system correlates successive time samples on the input to *estimate* the degree of congruence. We can talk of autocorrelation only when the successive time samples in fact contribute to the growth of $S(\rho\rho)$ for a particular pattern—and hence to the pattern's perceived clarity, according to the model.

If one postulates that the initial correlation of pattern and receptive field units is for filtering or feature extraction, the subsequent correlation of the units' outputs with themselves over time for image maintenance, the initial clarity of the stabilized image can be accounted for. Fading occurs because units adapt rapidly when there is no variation in their input, and hence no successive samples on which to form correlations. This sort of argument is supported by the findings that (*a*) acuity improves in a stabilized image for times up to about one second, and in the same way as for a nonstabilized target (Keesey, 1960), and (*b*) for a well-stabilized image, fading starts typically at between one and two seconds (Barlow, 1963). This suggests that the "instantaneous" correlation on the static input in fact takes some time to be computed and starts to be degraded after about a second. It would be of some interest to know whether the repetition clarity effect can be obtained on stabilized images, or whether retinal image motion is a necessary condition of its occurrence.

From considering the processing of one element, one should proceed to the problem of processing for a whole pattern, which may consist of a concatenation of straight lines in various orientations, or of other features such as curved lines, surfaces, points where two or more lines intersect, etc. An immediate difficulty here is that virtually nothing is known about the coding of pattern elements other than straight lines. It may be, as I have suggested elsewhere (Dodwell, 1970, Ch. 4), that for other pattern elements, perceptual learning is a factor, in which case a more complicated account of the signal processing would be necessary. At least we can say that the present model is consistent with the stabilized image data in postulating a rather basic role for the processing of straight-line segments.

Processing of Horizontal and Vertical Elements

There is a great deal of evidence, both physiological and psychophysical, to show that in human pattern vision, horizontal- and vertical-line elements play special roles. Much of it is summarized by Taylor (1963) and by Dodwell (1970). It is perhaps surprising that Hubel and Wiesel have consistently reported that the relative frequencies of their oriented line detectors are essentially random with respect to orientation when mapped for static displays (Hubel & Wiesel, 1962, 1968). Given the greater acuity for lines in the horizontal and vertical orientations, the higher evoked cortical potentials for horizontal and vertical as opposed to diagonally oriented gratings (Maffei & Campbell, 1970), and similar findings, one might expect there to be a preponderance of horizontal and vertical detector elements.[2] An alternative possibility is that the correlations and autocorrelations are performed primarily on the horizontal and vertical components of image shift. This would be in harmony with the known high acuity in these orientations, with Andrews' (1967) demonstration of more selective tuning for horizontal and vertical detectors in human vision, and with the finding of Pettigrew et al. (1968) that at the more complex level of processing for moving slits in binocular correspondence (in the cat), there is indeed the expected anisotropy with respect to orientation in cortical detector units. Maffei and Campbell's data on evoked potentials, referred to previously, also fit well with this hypothesis.

Retinal Receptive Field Size and Amplitude of Eye Movements

If the suggested method of computing autocorrelations is to be plausible, one would need to show that receptive field sizes and the amplitude of eye movements are of the right order of magnitude for the generation of such functions. There is no firm evidence on receptive field size in the human visual system, but Andrews (1967) estimates that fields for line detector units with foveal representation are about nine minutes of arc long. No one knows how closely packed the units may be, or how elongated, but a reasonable guess would be that such fields would be about one minute of arc across, so that the central region of a line detector might be between 10 seconds and 30 seconds across. The average amplitude of nystagmus is around 15 seconds to 20 seconds, a value which is large enough to ensure that pattern elements would be shifted from "on" to "off" regions of single receptive fields, and since fields almost certainly overlap a good deal, to shift maximum firing rates from unit to unit within a column. In this connection, the fact that there is a sudden reduction in fading when image motions on the retina with amplitudes over about 20 seconds are allowed to occur (Ditchburn, Fender, & Mayne, 1959) looks significant. Also, Krauskopf (1957) showed that the lower frequency components of eye movements (under 10 per second) are most effective in maintaining target visibility. Thus it seems that, as far as it goes, the evidence is consistent with the possibility of autocorrelation generation within the receptive field contour coding system.

---

[2] Admittedly there is a jump across species in this inference; but it is not unreasonable to suppose that a binocular system such as the monkey's will have similar mappings to those in the human contour coding system, hence to infer randomness with respect to orientation for the latter also.

EVOLUTION OF THE SYSTEM

The proposed type of organization can be viewed as the result of developments in a structure subjected to two conflicting evolutionary pressures. The biological utility to an organism of improved acuity and capacity for detail vision is clear enough; so are the advantages of having eyes which can be moved rapidly and precisely to orient toward and fixate on visual features in widely separated directions from the present line of regard. The conflict arises because the former property would lead one to expect structures with large receptor surfaces to evolve, which could be maintained in a fixed and steady orientation; the second should lead to the development of small structures controlled by powerful, but precisely gradable, muscles. In the former case, large structures with a low power-to-weight ratio of extrinsic muscles to eye would be suitable. In the latter, a high power-to-weight ratio is desirable, but with several strong muscle systems controlling the eye simultaneously, and under antagonistic tensions, fine tremors of the structure are inevitable. The tremors could be damped out by increasing the moment of inertia of the eye, but this would defeat the whole purpose of the system. Development of processes whereby the muscle tremors themselves became an integral part of the signal extraction system for contour details presents an elegant compromise between the two conflicting specifications for an efficient visual apparatus described above. From this point of view, the proposed model makes good sense biologically.

Although it would be difficult to argue that the human visual system evolved precisely to handle the sorts of information processing to which experimental psychologists subject it, one can argue that the limitations on storage and sequential processing (other than sequential spatial scanning which appears to develop from reading habits) are indeed results of evolutionary pressures very similar to those just discussed. A time-integration system is efficient for detection of weak visual signals, but time lag is also inefficient from the point of view of processing of rapidly changing inputs (the problem presented at the beginning of this paper). In the following section it is shown that a correlational model solves the problem of these conflicting desiderata too.

CLARITY, STORAGE, AND INFORMATION PROCESSING

Convincing evidence for a short-term visual store was provided by Sperling (1960) using a partial-report technique for arrays of letters; other investigators have since found evidence for such a process under a wide variety of conditions (e.g., Averbach & Coriell, 1961; Dodwell & Engel, 1963; Keele & Chase, 1967; Eriksen & Collins, 1967; Standing & Dodwell, in press; Standing, Haber, Cataldo, & Sales, 1969). Most of the evidence is consistent with the idea that even for very brief stimulus durations, short-term visual storage lasts for 200 to 300 milliseconds, the amount of persistence depending to some extent on other exposure conditions and the observer's state of adaptation. As far as the autocorrelation interpretation is concerned, two points should be made: first, it has been shown by direct measures of short-term visual storage that the actual duration of the stimulus exposure may be relatively unimportant; as long as the blank interval between repeated supraliminal presentations of a display is less than about 250 milliseconds, observers report that it does not fade completely (Haber & Standing, 1969). Second, summative effects between repetitions of subthreshold displays, that is to say, displays which are below recognition threshold for the information they contain, can occur over intervals as long as 900 milliseconds, when the image certainly fades completely between presentations (Standing et al., 1969). Taken together, these findings show that perceptual persistence is not simply a function of stimulus energy, and that useful information can be stored well beyond the time for phenomenal persistence (cf. also Daw, 1962). It is reasonable to propose that whatever is stored as a record of the spatial pattern information in the display is stored in a form that is *abstracted* from the input display, and that $S(\rho\rho)$ or some close analogue of it would be a plausible candidate. There are some problems with this idea, however: $S(\rho\rho)$ is simply a number or quantity which expresses a property of a pattern, but does not reflect any of the specific pattern features. The autocorrelation function itself contains at least a record of the spatial frequencies in a pattern, so that it or its Fourier transform could be the storage medium. Such a notion is reinforced by the fact that there is accumulating evidence to show that at least under certain conditions the visual system operates as a spatial frequency analyzer (e.g., Campbell & Kulikowski, 1966; Julesz & Stromeyer, 1970).

There is evidence that random noise fields and other masks interfere with information processing in the visual system by stopping processing of earlier presented material rather than by summating with a given input (Averbach & Coriell, 1961; Haber & Standing, 1968); a good criterion for deciding whether or not successive rapid inputs to the visual system are spatial transformations of the same pattern, to be processed together as equivalent displays, would be a criterion which preserves some informational features which are invariant under translation on the array; correlation between successive time samples on the visual input seem to fit the bill. Where the same pattern is present in successive time samples, an autocorrelation function would be built up; where different patterns follow each other in sequence, mutual interference or partial integration would occur depending on the patterns' similarities which in turn are estimated in terms of cross-correlations between successive time samples. Development of this notion might provide some new insights into the effectiveness of different sorts of visual mask, and resolution of the controversy about the principle mechanisms of masking (Kahneman, 1968). It can be applied also to the recognition of characters in dynamic visual noise (Uttal, 1969a, 1969b).

The idea of time sampling is supported by evidence for the quantization of perceptual time (Dodwell & Standing, 1970; Kristofferson, 1967), and measures of the size of the perceptual moment should tell us something about the time required to form cross-correlations and autocorrelations. On the basis of simultaneity judgments and masking of single flash stimuli, 40–50 milliseconds appears to be the right order of magnitude for this. However, for more complex arrays, longer integration periods would be probable, so that 40 milliseconds is a lower bound for this type of processing. Evidence on integration of pattern information in letter displays, for example, provide estimates of at least 100 milliseconds (Eriksen & Collins, 1967).

At a somewhat greater remove, it is worth noting that performing autocorrelations on time series to extract signal characteristics from noise has long been a standard technique in communications research, and has obvious possibilities in audition, for example, in the detection of pure tones masked by noise. The finding that short-term memory for letters and numbers presented visually seems to be echoic in character (Conrad, 1964) suggests again the possibility of low-level signal processing and storage in terms of the physical features of sensory input, in this case with some form of cross-modality coupling; perhaps "covert rehearsal" of auditory material is a form of autocorrelation. However, these notions are at best suggestive. To give substance to the model, one must show that some fairly specific predictions follow from it, predictions which do not follow from other possible models for perceptual clarity, information processing, and storage.

SOME PREDICTIONS FROM THE MODEL

The predictions to be made rely heavily on the assumption that autocorrelations are performed primarily on the horizontal and vertical components of visual inputs in the manner described earlier and that because of the physiological mechanism involved, finite amounts of time are needed to form the autocorrelation.

1. The rate of growth of clarity should be a decreasing monotonic function of pattern complexity, where complexity is described by the amounts of contour in a display and their horizontal and vertical spatial frequencies. Studies of the repetition clarity effect to date have all used letter displays,

and complexity as here defined has not been manipulated explicitly. The prediction is that for very "simple" patterns such as a line grating (where a single horizontal or vertical spatial frequency is present), rate of growth in clarity should be very rapid, whereas for a random grid, or a random dot pattern such as a Julesz field, it should be slow. Some informal observations on Julesz fields (Psotka, personal communication, 1970) indicate that growth of clarity is indeed very slow for such displays.

2. For a given display, such as a letter or a sequence of letters, the rate of growth of clarity should depend on the particular type font used: a format which consisted mainly of horizontal and vertical contours should both show greater initial clarity and achieve an asymptotic level more rapidly than one which consisted of curves and oblique lines and contained numbers of irrelevant features. As a corollary, rate of growth of clarity for handwritten script should be lower than for machine-typed (or hand-typed) letters.

3. Variations in type font from presentation to presentation should slow down the growth of clarity. There is one piece of evidence which goes against this prediction (Haber, Standing, & Boss, 1970); however, the curve published by these authors (Figure 1 of their paper) suggests that when typeface and position of letters are varied, the rate of growth in clarity is slowed down, although the effect was not reliable in their experiment. Earhard and Fullerton (1969) have shown that under certain rather restrictive conditions, variation in line width in printed letters does not become easier to detect over repeated trials, although individual letters do become clearer. A more thorough test of the hypothesis seems warranted.

4. Introduction of visual noise into a display should slow down growth in clarity, as should introduction of irrelevant displays between presentations of the target display.

5. In a discrimination paradigm, where two different displays have to be distinguished from each other in repeated presentations, some interference with the growth of clarity might be expected. However, this prediction cannot be made with conviction, since one does not know to what extent information about independent but concurrent displays can be stored simultaneously without mutual interference. This question would be worth investigating in its own right.

6. Following 5, if *different* noise fields are used to mask successive presentations of the *same* target, the effectiveness of the mask should decline over trials. This could in fact constitute a sensitive test of the extent to which masking actually stops visual processing.

7. The effectiveness of a mask with respect to a given target should depend on their structural similarities. This is a very vague statement; in fact the systematic investigation of mask-probe features should be a good way of determining what the dimensions of structural similarity are. As a first guess, one can propose that the most effective masking should occur when probe and mask both contain several spatial frequencies, but share none in common.

8. Stereoptic depth should grow in clarity on repeated presentations, and should be some function of the changes in clarity of the monocular displays. For example, the growth in clarity for stereo depth should be more rapid if either or both monocular displays have reached full clarity (by repeated single or alternating monocular exposures) than if dichoptic stimulation is used from the start of a series.

9. Reversals (up–down and left–right) of patterns—even complex ones containing nonhorizontal or vertical components—should interfere with growth in clarity much less than small rotations of the display. This follows from the fact that the postulated autocorrelation function extracts spatial frequency components along the horizontal and vertical axes, components which are not affected by reversals, but are changed by rotations. The prediction is restricted to the class of patterns which do not display complete identity under rotation, such as a set of concentric circles, obviously.

10. Some auditory analogues suggest themselves, for example, repetition clarity effects for tones and spoken material pre-

sented in noise. It would be of considerable interest to find out whether cross-modality effects on clarity and repetition could be obtained between vision and hearing.

A problem exists for measuring clarity effects in situations where parts of a display are not easily named, as in a random dot field. Individual letters in an array can be named, but obviously not individual dots, or bars in a grating. Some other response indicator would be needed; probably a form of scaling such as magnitude estimation (Stevens, 1957) would be suitable, so that the problem is not insuperable.

CONCLUSION

The predictions are not all unique to a model which specifies a form of correlation between consecutive time samples as the basis for perceptual clarity, storage, and interference between rapid input processes. However, it is difficult to think of another model which would encompass them all. Neither a Hebbian cell-assembly model, nor one that postulates the extraction of specific features from a display before synthesis of the to-be-recognized patterns, is inconsistent with the predictions, although such models do not lead explicitly to them. But those models are not inconsistent with the idea of autocorrelation as a basic operation in signal processing either. As was maintained earlier, the correlating of spatial patterns with themselves is to be thought of as a process which occurs "low down" in the system (although not peripherally, in the physiological sense) and, therefore, would be a means of determining what the distinctive features, or the inputs to a cell assembly, might be.

No one of the lines of evidence or arguments presented—on the repetition clarity effect, fading and regeneration of stabilized images, visual acuity, short-term visual storage and masking, and biological utility—is by itself convincing support for the model, but taken together they add up to a consistent and quite appealing account of how perceptual clarity can be explained.

The notion that autocorrelation plays a fundamental role in image processing and storage has gained wide currency in recent years,[3] the main emphasis having been on computer applications, as was earlier pointed out (e.g., Rosenfeld, 1969). Very recently, the idea has been extended in a general way to biological systems too (e.g., Anderson, 1969; Barrett, 1969). In its strongest form, the postulation of an autocorrelational system claims that the brain operates like a hologram (van Heerden, 1968; Longuet-Higgens, 1968; Pribram, 1966), which may be reasonable considering the enormous long-term storage capacity—particularly for pictorial material—of the normal human being. The present proposals are more modest in assigning to autocorrelations a specific function in visual processing. At the same time, they are probably more open to empirical test than the general holographic model. Tests on the specific model would, on the other hand, lend some support to the idea that autocorrelation plays a basic role in information processing in the central nervous system as a whole, to extracting signals from noise, to matching an input to a degraded stored version, to perceptual clarity.

---

[3] I am indebted to Joseph Psotka for pointing this out and for drawing the relevant literature to my attention.

REFERENCES

ANDERSEN, E. E., & WEYMOUTH, F. W. Visual perception and the retinal mosaic. I: Retinal mean local sign. *American Journal of Physiology*, 1923, **64**, 561–594.

ANDERSON, J. A. A memory storage model utilizing spatial correlation functions. *Kybernetik*, 1969, **5**, 113–119.

ANDREWS, D. P. Perception of contour orientation in the central fovea. Part II: Spatial integration. *Vision Research*, 1967, **7**, 999–1013.

AVERBACH, E., & CORIELL, A. S. Short-term memory in vision. *Bell System Technical Journal*, 1961, **40**, 309–328.

BARLOW, H. B. Slippage of contact lenses and other artifacts in relation to fading and regeneration in supposedly stable retinal images. *Quarterly Journal of Experimental Psychology*, 1963, **15**, 36–51.

BARLOW, H. B., BLAKEMORE, C., & PETTIGREW, J. D. The neural mechanism of binocular depth discrimination. *Journal of Physiology*, 1967, **193**, 327–342.

Barrett, T. W. The cerebral cortex as a diffractive medium. *Mathematical Biosciences,* 1969, **4,** 311–350.

Campbell, F. W., & Kulikowski, J. J. Orientational selectivity of the human visual system. *Journal of Physiology,* 1966, **187,** 437–445.

Conrad, R. Acoustic confusions in immediate memory. *British Journal of Psychology,* 1964, **55,** 75–86.

Daw, N. W. Why after-images are not seen in normal circumstances. *Nature,* 1962, **196,** 1143–1145.

Ditchburn, R. W., Fender, H. D., & Mayne, S. Vision with controlled movements of the retinal image. *Journal of Physiology,* 1959, **145,** 98–107

Ditchburn, R. W., & Ginsborg, B. L. Vision with a stabilized retinal image. *Nature,* 1952, **170,** 36–37.

Dodwell, P. C. *Visual pattern recognition.* New York: Holt, Rinehart & Winston, 1970.

Dodwell, P. C., & Engel, G. R. A theory of binocular fusion. *Nature,* 1963, **198,** 39–40, 73–74.

Dodwell, P. C., & Standing, L. G. Studies of visual backward masking and a model for the Crawford effect. *Acta Psychologica,* 1970, **32,** 31–47.

Earhard, B., & Fullerton, R. How much does repetition facilitate perception? *Journal of Experimental Psychology,* 1969, **81,** 101–108.

Engel, G. R. The autocorrelation function and binocular brightness mixing. *Vision Research,* 1969, **9,** 1111–1130.

Ericksen, C. W., & Collins, J. F. Some temporal characteristics of visual pattern perception. *Journal of Experimental Psychology,* 1967, **74,** 476–484.

Evans, C. R. Some studies of pattern perception using a stabilized retinal image. *British Journal of Psychology,* 1965, **56,** 121–133.

Haber, R. N. Repetition as a determinant of perceptual recognition processes. In R. N. Haber (Ed.), *Information processing approaches to visual perception.* New York: Holt, Rinehart & Winston, 1969.

Haber, R. N., & Hershenson, M. Effects of repeated brief exposures on the growth of a percept. *Journal of Experimental Psychology,* 1965, **69,** 40–46.

Haber, R. N., & Standing, L. G. Clarity and recognition of masked and degraded stimuli. *Psychonomic Science,* 1968, **13,** 83–84.

Haber, R. N., & Standing, L. G. Direct measures of short-term visual storage. *Quarterly Journal of Experimental Psychology,* 1969, **21,** 43–54.

Haber, R. N., Standing, L. G., & Boss, J. Effect of position and typeface variation on perceptual clarity. *Psychonomic Science,* 1970, **18,** 91–92.

Heckenmueller, E. G. Stabilization of the retinal image: A review of method, effects, and theory. *Psychological Bulletin,* 1965, **63,** 157–169.

Hershenson, M., & Haber, R. N. The role of meaning in the perception of briefly exposed words. *Canadian Journal of Psychology,* 1965, **19,** 42–46.

Hubel, D. H., & Wiesel, T. N. Receptive fields, binocular interaction, and functional architecture in the cat's visual cortex. *Journal of Physiology,* 1962, **160,** 106–154.

Hubel, D. H., & Wiesel, T. N. Receptive fields and functional architecture in monkey striate cortex. *Journal of Physiology,* 1968, **195,** 215–243.

Julesz, B., & Stromeyer, C. F. Masking of spatial gratings by filtered one-dimensional noise. Paper presented at the meeting of the Psychonomic Society, San Antonio, Texas, November 1970.

Kahneman, D. Method, findings, and theory in studies of visual masking. *Psychological Bulletin,* 1968, **65,** 404–425.

Keele, S. W., & Chase, W. G. Short-term storage. *Perception and Psychophysics,* 1967, **2,** 383–386.

Keesey, V. K. Effects of involuntary eye movements on visual acuity. *Journal of the Optical Society of America,* 1960, **50,** 769–774.

Krauskopf, J. Effect of retinal image motion on contrast thresholds for maintained vision. *Journal of the Optical Society of America,* 1957, **47,** 740–747.

Krauskopf, J., Cornsweet, T. N., & Riggs, L. A. Analysis of eye movements during monocular and binocular fixation. *Journal of the Optical Society of America,* 1960, **50,** 572–578.

Kristofferson, A. B. Attention and psychological time. *Acta Psychologica,* 1967, **27,** 93–100.

Longuet-Higgens, H. C. Holographic model of temporal recall. *Nature,* 1968, **217,** 104–106.

MacLachlan, D., Jr. The role of optics in applying correlation functions to pattern recognition. *Journal of the Optical Society of America,* 1962, **52,** 454–459.

Maffei, L., & Campbell, F. W. Neurophysiological localization of the vertical and horizontal visual coordinates in man. *Science,* 1970, **167,** 386–387.

McKinney, J. P. Verbal meaning and perceptual stability. *Canadian Journal of Psychology,* 1966, **20,** 237–242.

Pettigrew, J. D., Nikara, T., & Bishop, P. O. Binocular interaction of single units in cat striate cortex: Simultaneous stimulation by single moving slit with receptive fields in correspondence. *Experimental Brain Research,* 1968, **6,** 391–410.

Pribram, K. H. Some dimensions of remembering: Steps toward a neuropsychological model of memory. In J. Gaito (Ed.), *Macromolecules and behavior.* New York: Academic Press, 1966.

Pritchard, R. M., Heron, W., & Hebb, D. O. Visual perception approached by the method of stabilized images. *Canadian Journal of Psychology,* 1960, **14,** 67–77.

RATLIFF, F., & RIGGS, L. A. Involuntary motions of the eye during monocular fixation. *Journal of Experimental Psychology,* 1950, **40,** 687–700.

RIGGS, L. A., RATLIFF, F., CORNSWEET, J. C., & CORNSWEET, T. N. The disappearance of steadily fixated visual test objects. *Journal of the Optical Society of America,* 1953, **43,** 495–501.

ROSENFELD, A. *Picture processing by computer.* New York: Academic Press, 1969.

SHORTESS, G. K., & KRAUSKOPF, J. Role of involuntary eye movements in stereoscopic acuity. *Journal of the Optical Society of America,* 1961, **51,** 555–559.

SPERLING, G. The information available in brief visual presentations. *Psychological Monographs,* 1960, **74**(11, Whole No. 498).

STANDING, L. G., & DODWELL, P. C. Retroactive enhancement: A new visual storage effect. *Quarterly Journal of Experimental Psychology,* in press.

STANDING, L. G., HABER, R. N., CATALDO, M., & SALES, B. D. Two types of short-term visual storage. *Perception and Psychophysics,* 1969, **5,** 193–196.

STEVENS, S. S. On the psychophysical law. *Psychological Review,* 1957, **64,** 153–181.

TAYLOR, M. M. Visual discrimination and orientation. *Journal of the Optical Society of America,* 1963, **53,** 763–765.

UTTAL, W. R. Masking of alphabetic character recognition by dynamic visual noise (DVN). *Perception and Psychophysics,* 1969, **6,** 121–128.

UTTAL, W. R. The character in the hole experiment: Interaction of forward and backward masking of alphabetic character recognition by dynamic visual noise (DVN). *Perception and Psychophysics,* 1969, **6,** 177–181.

VAN HEERDEN, P. J. *Foundations of empirical knowledge.* Wassenaar, Netherlands: M. V. Uitgererij Wistik, 1968.

VON HOLST, E. Relations between the central nervous system and the peripheral organs. *British Journal of Animal Behaviour,* 1954, **2,** 89–94.

VON HOLST, E., & MITTELSTAEDT, H. Das Reafferenzprinzip. *Die Naturwissenschaften,* 1950, **37,** 464–476.

(Received August 24, 1970; revision received January 13, 1971)

From July 15, 1971, to July 15, 1972, manuscripts should be submitted to the Editor, George Mandler, *Psychological Review,* Department of Experimental Psychology, South Parks Road, Oxford OX1 3PS, England. With the location of the editorial office in England during that year, it is especially important that authors submit their manuscripts *in triplicate* and pay special attention to APA style requirements.

Psychological Review
1971, Vol. 78, No. 4, 290–302

ORGANIZATION AND INTERFERENCE[1]

LEO POSTMAN[2]

University of California, Berkeley

This paper presents evidence for the functional continuity of the products of subjective organization and of associative learning. A series of studies is reported in which the transfer effects of free recall on paired-associate learning were investigated. Under the paradigm used in these studies, the words in the successive lists remained the same for the experimental groups and were unrelated for the control groups. The results consistently showed negative transfer in the acquisition of the paired-associate list. Evidence was also obtained for long-term proactive inhibition in the recall of the transfer task. Thus, the higher order units developed in the free recall phase were an effective source of interference when the transfer task required the establishment of new relations among the component items. The interitem dependencies established in free recall and in associative learning are viewed as commensurable phenomena that are likely to be governed by common principles. The implications of these results for the analysis of extraexperimental sources of interference are discussed.

In contemporary analyses of learning and retention, the concepts of association and of organization appear to stand in apparent opposition to each other. These concepts have their origins in different theoretical traditions and are characteristically anchored to different experimental operations. The classical paradigm for the analysis of associative processes has been, as the very name implies, the paired-associate task. The acquisition of the prescribed pairings signals the establishment of associative connections. The primary vehicle for the examination of organizational processes has been the free recall experiment. The development of higher order units is inferred from the increasing consistency in the order of output as learning progresses (Tulving, 1962b). The differences in experimental operations and methods of measurement should not discourage us, however, from exploring systematically the continuity of the phenomena studied in these paradigmatic situations. To be successful, such an exploration will undoubtedly require, as Tulving (1968) put it, a "liberalization of the orienting attitudes that lie at the root of traditional formulations [p. 6]."

It is the thesis of this paper that the subjective units of free recall and the products of controlled associative learning are commensurable dispositions sharing common functional properties. To the extent that this thesis is valid, the question of what is learned should not be a source of basic disagreement between exponents of association and of organization. Rather, the two positions will be seen as facing common explanatory problems. This is not to argue for the primacy of the principles of association or, for that matter, of organization. Given agreement on what is learned, the question still remains open as to which formulation has greater analytic and predictive power.

DEVELOPMENT OF INTERITEM DEPENDENCIES

The products of free recall learning and of controlled associative practice can be readily specified within a common descriptive framework. In both cases, the results of training are reflected in the increased interdependence of nominal units: Given A, either B or C or D, etc., may be expected to follow with varying probabilities. This basic descriptive statement can be readily modified to accommodate extended sequences or even hierarchical trees of categorized items (Bower, Clark, Lesgold, & Winzenz, 1969). It also encompasses the limiting case in which the probability is unity for (say) B and zero for all other alternatives. The measurable differences between subject-generated and experimentally prescribed sequences then reduce to characteristics of such probability distributions. In the former case, the groupings are optional and freely expandable; in the latter case, they are prescribed and fixed. Thus, in free recall the definition of the task for the subject permits the development of cumulative groupings that may be more or less rigid in their internal ordering. The experimental arrangements and conditions of testing in paired-associate learning, on the other hand, impose a pairwise module on interitem dependencies.

---

[1] This paper represents a modified version of the writer's Presidential Address to Division 3 of the American Psychological Association, Miami Beach, September 5, 1970. The research reported in the paper was supported by Grant MH 12006 from the National Institute of Mental Health. Thanks are due to Janat Parker and Sheila Burns for their assistance.

[2] Requests for reprints should be sent to Leo Postman, Institute of Human Learning, University of California, Berkeley, California 94720.

A difference in the conditions of practice and in the criteria of acceptable performance does not, of course, imply a fundamental disparity of underlying processes. For example, all the available evidence indicates that the processes of intentional and incidental learning are governed by the same principles (cf. Postman, 1964). There may, in fact, be a useful parallel here to the distinction with which we are concerned: What is selected incidentally is optional, what is practiced intentionally is prescribed. Nevertheless, the same functional relations appear to obtain for both. There is no compelling reason to believe that there is any greater discontinuity between the subjective units developed in free recall and the products of controlled associative learning.

The apparent opposition between the concepts of organization and of association in the face of commensurable facts appears to stem in part from the divergent emphases that the two positions place on the factor of contiguity. Traditionally, association theory has regarded contiguity of stimulus and response as a necessary condition of learning. This assumption is reflected in the preference for such classical paradigms as paired-associate and serial learning where contiguity is built into the experimental arrangements. In modern theoretical analyses of association, however, contiguity has become a formal construct which is only precariously anchored to experimentally controlled conditions of physical or temporal proximity. We need to think only of the significance attached to mediators in the analysis of associative processes. Here the critical contiguities are between observed and implicit events. Yet the silent presence of the schema of literal contiguity has probably dampened the enthusiasm of association theorists for the intensive exploration of the phenomena of subjective organization. There have been notable exceptions such as Deese's (1959, 1960) work. His studies have the clear implication that interitem associations can generate organization. There is no conceptual barrier to an associationistic analysis of higher order groupings that develop among items whose contiguity relations change from trial to trial.

Organization theorists have been preoccupied with another type of contiguity, namely, that between particular responses in successive tests of free recall. In the interpretation of output regularities that develop in spite of changing orders of input, concepts other than association, such as chunking and unitization, have been favored. Whatever the preferred label, output consistency is not likely to prove a sensitive gauge of organization in the broad sense of the term, that is, if organization is taken to denote the development of systematic interitem dependencies. The relationship between organization and output consistency will be seen as uncertain and fallible if it is granted that interitem dependency need not be reflected in rigid response sequences. There is, in fact, a convincing argument to the contrary. As a multiplicity of interitem relations develops, an increasing number of equivalent pathways leading from one response to another becomes available. Thus, cohesiveness of grouping may imply flexibility of access. It is only when organization is based on rigidly ordered categories, such as in the use of an alphabetical sequence (e.g., Tulving, 1962a), that intertrial repetition is likely to mirror faithfully the development of the schema. To speak of rote organization may not be a contradiction in terms!

It is evidence of interitem dependencies, then, that constitutes the empirical basis for inferences about both associative and organizational processes. For purposes of associationistic analysis, controlled contiguity of input is not an essential requirement; the critical contiguities may be mediated by implicit events during the processing of the input. In the application of the concept of organization, the degree of contiguity of responses in output probably does not reflect in a one-to-one fashion the amount of structure imposed on the materials. Depending on the characteristics of the materials and the constraints of the task requirements, the groupings that develop may range from fixed pairwise linkages to a complex network, or an aggregate of networks such as is found in categorized lists. The mechanisms responsible for the establishment of interitem dependencies—the development of what Underwood (1969) recently described as the associative attributes of memory essential for retrieval—remain to be fully understood.

SUBJECTIVE ORGANIZATION AND ASSOCIATIVE INTERFERENCE

The argument for continuity of inferred associative and organizational processes cannot rest on a priori considerations alone but must be documented experimentally. A basic empirical question that grows out of the present analysis is whether subjective organization in free recall entails the development of linkages between discrete units that share the functional properties of the sequences established in controlled associative learning. In support of an affirmative answer, it is not sufficient to point to the regularities in the order of output that are conventionally used as an index of organization. There remains some uncertainty as to whether the recurrent response sequences reflect the acquisition of higher order units or merely a systematic ordering of available items on successive recall trials. It has, in fact, been argued that there is independence of traces in storage and that organization is first imposed on recall at the time of test by the adoption of a retrieval plan (Slamecka, 1968, 1969). The argument will be difficult to resolve as long as the evidence for organization is sought in internal analyses of free recall performance. Here, as in other cases, a more promising approach to the assessment of what has been learned is provided by tests of transfer. If subjective groupings are a product of learning (or a property of storage), it follows that after free recall practice there should be interference in the acquisition and retention of transfer tasks that require the disruption of such groupings. The logic of this approach starts with the strong assumption that progress in free recall learning implies ipso facto the achievement of organization. Once such organization has occurred, it will become a source of interference when the transfer task requires the establishment of new relations among the component items.

The basic operations required for a test of this prediction are those of a two-stage transfer experiment. In the first stage, a list of words is learned by the method of free recall and an opportunity is provided for the development of higher order groupings. In the second stage, a list of paired associates is learned in which both the stimulus and the response terms are drawn from the prior free recall task. The arbitrary pairings of the items in the transfer stage should be in conflict with the groupings developed in the course of free recall learning. Consequently, there should be negative transfer in the acquisition of the paired-associate list relative to a control condition in which the materials learned in the first and in the second stage are unrelated. This is a strong prediction because the associative interference is expected to outweigh the positive effects of item familiarization attendant on the prior free recall learning of the stimuli and the responses. The original organizational linkages will eventually be supplanted during the mastery of the transfer task. To determine whether they recover to become sources of proactive interference, a third stage must be added to the experiment: the measurement of long-term retention for the paired-associate task. The question is once more whether organizational linkages function like prescribed associations, this time in the context of a classical paradigm of interference at recall.

The consideration of long-term proactive losses calls attention to potential implications of our prototype experiment for the analysis of preexperimental sources of interference in forgetting. The subjective groupings, which are developed during the free recall phase of the experiment and which may recover during the retention interval, will be based on the subject's preexperimental verbal habits. The sequence of tasks may be viewed, therefore, as tapping, at least indirectly, preexperimental sources of interference in the recall of new associations acquired in the laboratory. It is, of course, a deplorable but indisputable fact that evidence for the influence of prior language habits on retention has been notoriously difficult to obtain (cf. Ekstrand & Underwood, 1965; Postman, 1963). One possible reason for the failure of the interference hypothesis in conventional studies of single-list retention may be the discontinuity between the laboratory task and the subject's normal verbal activities. Consequently, the learner's preexperimental habits are sharply differentiated from the associations prescribed in the laboratory. If free recall learning does, indeed, entail the development of higher order units based on preexperimental verbal dispositions, then the present design should favor the arousal and subsequent recovery of the critical linguistic habits in the context of the experimental situation.

EXPERIMENT I

The first study tested the hypothesis that the organization developed during the free recall learning of a list is a source of associative interference when items from that list are paired arbitrarily in the transfer stage.

Method

Design. This study comprised two experiments which were run concurrently. In each experiment free recall learning of a 20-word list was followed by the acquisition of a list of 10 paired associates. Under the experimental conditions, the same words were used in free recall and in paired-associate learning; under the control conditions, the words in the two lists were unrelated. The two experiments differed only with respect to the method of practice in the paired-associate phase. In one case, the transfer list was learned by the recall method, and in the other case, by the multiple-choice method. Performance in the paired-associate phase was likely to be influenced not only by associative transfer but also by prior familiarization with the response terms. Hence the multiple-choice procedure, which eliminates the requirement of response recall, was used to obtain as pure an estimate as possible of the associative component of transfer.

Under both methods of practice, additional experimental and control groups were included for which the order of the tasks was reversed, that is, the acquisition of the paired-associate list preceded free recall learning. The purpose of this expansion of the design was to demonstrate that pairwise association and item organization in free recall learning are not necessarily incompatible.[3] On the contrary, preexisting linkages should facilitate the development of higher order groupings in free recall; if so, there would be positive transfer from paired-associate to free recall learning. In summary, then, the experiments investigated transfer between free recall and paired-associate learning for both the possible task sequences, with the pool of items remaining the same under the experimental treatments and changing under the control treatments.

Lists. The learning materials were four lists of 20 two-syllable nouns of high to medium frequency of usage. These lists were used equally often in the free recall phase. Four different lists of 10 paired associates, with the pairings selected at random, were derived from each basic 20-word list. Thus, there were 16 paired-associate lists which were used under all conditions. The four basic lists were divided into two sets for purposes of assignment to the control treatments. Normative associative relations were minimized within lists and between the successive lists learned by control *S*s.

Procedure. In the free-recall phase, the words were presented at a 1.5-second rate in the window of a memory drum. The items appeared in different random orders on successive study trials. Each presentation of the list was followed by a 1.5-minute written test of free recall. When free recall learning was the first task, acquisition was carried to a criterion of 16/20 plus three additional study–test cycles.

In both the recall and the multiple-choice methods of paired-associate learning, study and test trials were alternated. The pairs were presented in four different orders on the study trials, and four other orders were used on the test trials. The rate of presentation was 1.5 seconds on study trials, and 3 seconds on test trials. The interval between a study and a test trial was 3 seconds, and that between a test trial and the next study trial was 1.5 seconds. The two methods were dis-

---

[3] For other relevant studies of transfer from paired-associate learning to free recall, see Segal and Mandler (1967) and Wood (1969b, 1970).

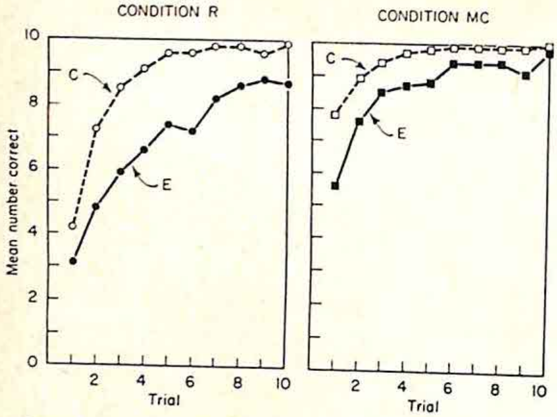


FIG. 1. Acquisition curves for paired-associate transfer lists in Experiment I.

tinguished by the procedure on the test trials. Under the recall method, each stimulus term was exposed for 3 seconds and the subject was required to supply the correct response. On multiple-choice test trials, four alternative responses were presented with each stimulus, and the subject had to call out the correct one during the 3-second exposure interval. All alternatives in the test items were responses from within the list. The set of incorrect alternatives accompanying a given correct response changed from one test order to the next. The subjects were instructed to respond to each test item, guessing if necessary. When the paired-associate list was learned first, acquisition under both methods was carried to the criterion of 8/10 plus three additional trials.

The transfer task, regardless of whether it was a free recall or a paired-associate list, was always practiced for 10 trials. The interval between the first and the second list was 3 minutes.

Subjects. There were 16 subjects in each of the eight groups. The subjects were undergraduate students at the University of California who were not necessarily naive to verbal learning experiments. The assignment to conditions was in blocks of eight, with one subject from each condition per block. The running order within blocks was determined by means of a table of random numbers.

Results

First-list learning. When the first task was free recall learning, the mean number of trials to the criterion of 16/20 was 4.55. For the initial paired-associate task, the mean number of trials to the criterion of 8/10 was 5.38 under the recall method and 2.78 under the multiple-choice method.[4] This difference was significant ($t = 3.31$, $df = 62$, $p < .01$). There were no reliable differences between the groups treated alike in the first phase.

Transfer. We turn now to the measures of performance in the test stage. Transfer from free recall to paired-associate learning will be considered first. In Figure 1, the acquisition curves for the recall method are shown on the left, and those for the multiple-choice method on the right. Under both procedures, the control group surpassed the experimental group. The differences were substantial and persistent. In an analysis of the total numbers of correct responses during second-list learning, the main effect of the condition of transfer was significant ($F = 11.81$, $df = 1/60$, $p < .01$). Acquisition was faster under the multiple-choice than under the recall method ($F = 10.32$, $df = 1/60$, $p < .01$). However, the negative transfer effects were clearly in evidence in each case. There was no reliable interaction between the method of testing and the condition of transfer. The contribution of response familiarization to transfer performance can be gauged from a comparison of the two sets of curves. For the recall method, the amount of negative transfer increased over several trials of test-list learning; the positive component of response learning initially masked some of the associative interference that was present. For the multiple-choice method, on the other hand, the separation between the experimental and the control group was maximal on the very first trial. The consistency of the results obtained by the two methods strongly supports the inference that prior list organization was a source of associative interference.[5]

An analysis of overt errors under the recall procedure yielded results entirely consistent with the findings on acquisition. As Table 1 shows, the mean percentages of intralist errors were higher under the experimental than under the control treatment

---

[4] In a previous study (Postman & Stark, 1969), in which the criterion was one errorless trial, no significant difference in first-list learning was found between the recall and the multiple-choice method. It appears that the advantage of the latter is limited to the early stage of acquisition.

[5] Wood's (1969a) finding of negative transfer from free recall to serial learning is consistent with this conclusion.

TABLE 1

MEAN PERCENTAGES OF INTRALIST ERRORS DURING TRANSFER LEARNING BY THE RECALL METHOD IN EXPERIMENT I

Condition	Misplaced responses	Stimulus intrusions	Total
Experimental	10.1	5.5	15.6
Control	1.7	1.6	3.3

($t = 2.30$, $df = 30$, $p < .05$) (after arcsine transformation). The separate rates of misplaced responses and intrusions of stimulus terms are also shown. The values were higher for the experimental than for the control group in both cases. The differences in the percentages of stimulus intrusions are particularly diagnostic. The prior occurrence of all items in a single free recall list impeded the differentiation between the stimulus and the response terms in the paired-associate task. Such failures of differentiation probably served to depress the speed of learning. They cannot, however, be regarded as an essential condition of negative transfer since they could not occur under the multiple-choice procedure. In the latter case the deficit in performance reflected directly the component of associative interference.

Transfer from paired-associate to free recall learning will be considered next. The acquisition curves for the test list are shown in Figure 2. Under both methods of practice, the experimental group was clearly superior to the control group. The separation was large on the initial trials but diminished as test-list learning continued and the ceiling of performance was approached. There was a slight reversal in the terminal stage of acquisition, particularly for the multiple-choice groups. The overall difference between the two treatments in the total number of correct responses was significant ($F = 12.79$, $df = 1/60$, $p < .01$). There was again no reliable interaction of the method of practice with the condition of transfer. The prior establishment of pairwise associations facilitated free recall learning but probably did not make for optimal list organization in the long run.

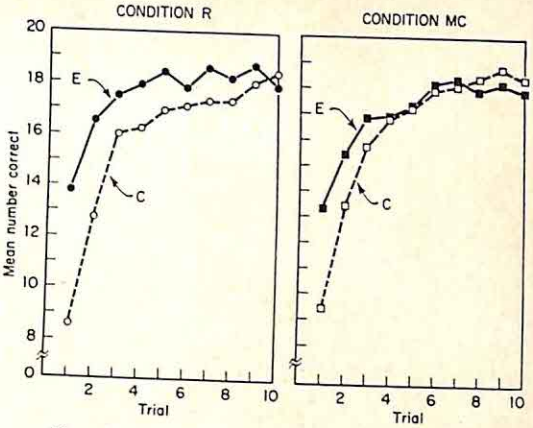


FIG. 2. Acquisition curves for free recall transfer lists in Experiment I.

There is no question but that the prior associations were utilized by the experimental subjects in free recall. The pairs from the previously learned list appeared as clusters in the subjects' output throughout free recall learning. The number of clusters in excess of chance (cf. Bousfield & Bousfield, 1966) increased as a function of trials, tracking the overall level of performance (Figure 3). For the control subjects, the frequency of occurrence of the same pairs remained, of course, close to zero throughout. It is interesting to note that the level of clustering was higher after recall than after multiple-choice training. The common mode of responding—active reproduction—encouraged the continued use of the same retrieval cues.

The critical finding is the asymmetry of the transfer effects between free recall and paired-associate learning: there was interference when free recall occurred first, and

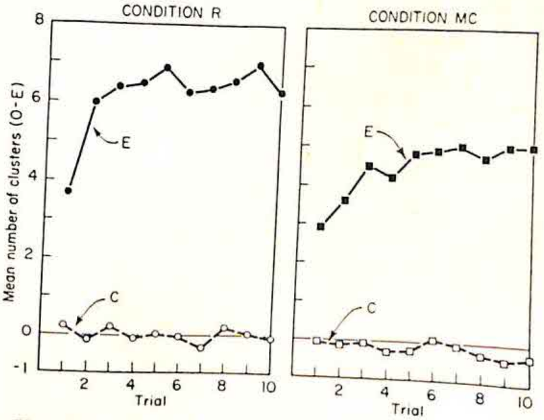


FIG. 3. Mean cluster scores for previously learned pairs during acquisition of free recall transfer lists in Experiment I.

facilitation when it occurred second. While the products of practice were carried over from one task to the other in each case, the sign of the transfer effects depended on whether or not the interitem relations established in the first phase remained appropriate in the second phase.

EXPERIMENT II

The next study continued the investigation of prior list organization as a source of interference. Attention was focused on the development of proactive inhibition as well as on negative transfer in acquisition.

Method

Design. The prototype design—free recall learning followed by the acquisition of a paired-associate list—was extended in two ways. First, in order to vary the potential degree of organization in free recall, both categorized and noncategorized lists were used as learning materials. Second, in order to assess the proactive influence of prior free recall learning on the retention of the transfer task, tests of retention for the paired-associate list were added. These tests were given either 1.5 minutes or 48 hours after the end of learning. Thus, the total design was a $2 \times 2 \times 2$ factorial, with 2 types of materials (categorized vs. noncategorized), 2 conditions of transfer (experimental vs. control), and 2 retention intervals (1.5 minutes vs. 48 hours). The conditions of transfer were defined as before: The items used in free recall and in paired-associate learning were the same under the experimental treatment and were unrelated under the control treatment.

Lists. The learning materials were six lists of 20 two-syllable nouns, four noncategorized and two categorized. The noncategorized lists were the same as those used in Experiment I. The categorized lists comprised four instances each of five taxonomic categories. Three of the categories were exhaustive (e.g., the four seasons of the year) and two were nonexhaustive (e.g., names of countries). In the construction of the paired-associate lists, noncategorized items were paired randomly, as was done in the previous experiment. In the case of the categorized materials, half the items from a given category appeared as stimuli and half as responses in the paired-associate list. Members of the same category were never paired. Four different sets of paired associates were derived from each 20-word list. Thus, there were 16 lists composed of noncategorized items and 8 composed of categorized ones. Lists were fully balanced over the appropriate portions of the design.

Procedure. The procedure in the free recall phase was exactly the same as in Experiment I; acquisition was again carried to the criterion of 16/20 plus three additional study–test cycles. The learning instructions made no reference to interitem relations and were exactly the same for the noncategorized and the categorized lists. In the latter case, the items from the different categories were presented in random order.

The paired-associate list was learned to the criterion of 8/10 by the recall method; the procedure was identical with that under the comparable conditions in Experiment I. The test of retention consisted of one recall trial without feedback, followed by five relearning trials.

Subjects. There was 16 subjects in each of the eight groups. The subjects were drawn from the same population as those in Experiment I and assigned to conditions in the same manner as before.

Results

First-list learning. The mean number of trials to the criterion of 16/20 was 4.22 for the noncategorized lists and 1.48 for the categorized lists. The difference between the two types of materials was significant ($t = 8.28$, $df = 126$, $p < .01$). There were no reliable differences between groups learning the same type of list.

The consistency of the order of output was, of course, much greater for the categorized than for the noncategorized material. The numbers of intertrial repetitions (Bousfield & Bousfield, 1966), in both the forward and the reverse order, were determined for the last three trials of acquisition. The mean score (obtained minus expected value) per trial was 6.13 for the categorized lists and 2.36 for the noncategorized lists ($t = 9.34$, $df = 126$, $p < .01$). In the former case, there was a high degree of clustering by categories. The mean number of categorical repetitions in excess of chance was 9.68 per trial. The high level of clustering indicates that the subjects readily discovered the categorized nature of the list. It can be assumed that the taxonomic groupings were well established when the subjects were shifted to the paired-associate phase.

Transfer. Figure 4 shows the acquisition curves for the transfer task. Successive criteria on the paired-associate list are plotted as a function of trials. It will be noted that the customary locations of the independent and the dependent variable are reversed in the figure. The resulting plot has the form of a conventional learning curve (cf. Hovland & Kurtz, 1952). This method of displaying the course of acquisition was chosen in order to permit a ready comparison with the transfer measures obtained in Experiment I (see Figure 1).

There are two main findings. First, the pairs composed of mismatched members of taxonomic categories were more difficult to learn than the pairs derived from a pool of unrelated words. The difference in trials to criterion was significant ($F = 32.30$, $df = 1/120$, $p < .01$). Second, for both types of list the test task was learned more slowly under the experimental than under the control condition ($F = 17.81$, $df = 1/120$, $p < .01$). The amount of negative transfer was somewhat greater for the categorized than for the noncategorized lists, but not reliably so. The results obtained with lists of unrelated words replicate the findings of the previous experiment. The presence of parallel effects for the categorized lists adds an important new dimension to our analysis of the conditions of associative interference—the degree of arousal of prior language habits in the experimental context.

On the face of it, no differences between the experimental and the control subjects learning the categorized test list might have been expected. The basic source of associative interference in the acquisition of that list was the mismatching of closely related items. The categorically appropriate response to each stimulus was present in the list and remained in sustained competition with the inappropriate prescribed one. The resulting interference stems from highly overlearned linguistic habits which both experimental and control subjects presumably brought to the laboratory. Nevertheless, there was a substantial amount of negative transfer. The interpretation of this fact is that the prior free recall practice served to activate the relevant linguistic habits in the experimental context and thus to increase their persistence and interfering potential in the transfer stage.

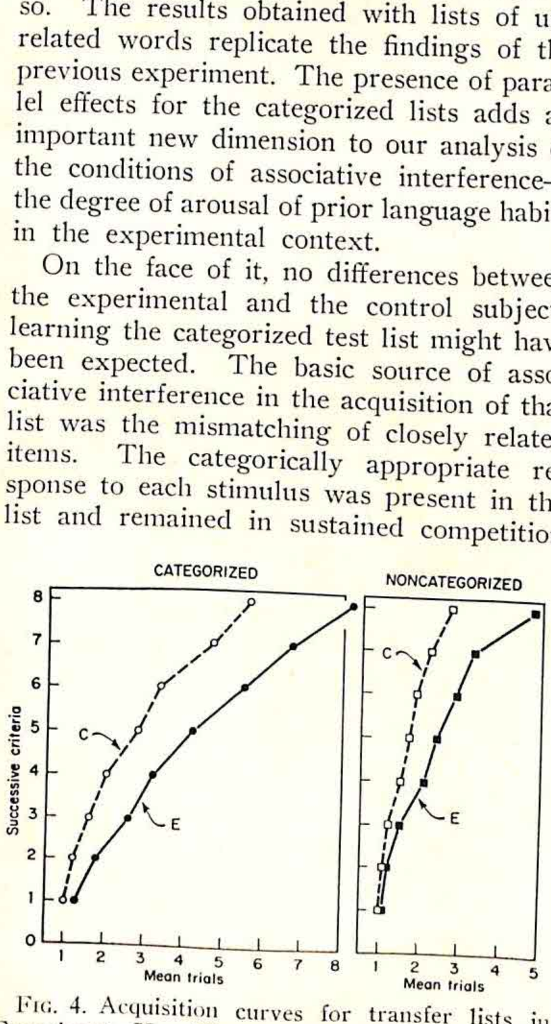


FIG. 4. Acquisition curves for transfer lists in Experiment II. (Successive criteria are plotted as a function of trials.)

TABLE 2

MEAN PERCENTAGES OF INTRALIST ERRORS DURING ACQUISITION OF CATEGORIZED AND NONCATEGORIZED TRANSFER LISTS: EXPERIMENT II

Condition	Misplaced responses	Stimulus intrusions	Total
Categorized list			
Experimental	12.7	12.9	25.6
Control	10.3	7.4	17.7
Noncategorized list			
Experimental	10.8	7.6	18.4
Control	4.6	2.3	6.9

As Table 2 shows, the percentages of intralist errors were higher for the experimental than for the control groups ($F = 20.63$, $df = 1/120$, $p < .01$); they were also greater for the categorized than for the noncategorized lists ($F = 17.93$, $df = 1/120$, $p < .01$). These differences obtained for both misplaced responses and stimulus intrusions. In the case of the categorized lists, "appropriate" intralist intrusions are of special interest, that is, the occurrence of erroneous responses which belonged to the same taxonomic category as the stimulus terms. The percentage of such intrusions, to the base of all overt responses, was reliably higher for the experimental than for the control group (3.1 vs. 1.8). However, relative to the total number of errors of commission, the frequencies of appropriate intrusions were comparable for the two groups (12 vs. 13%). Thus, the interference was more persistent under the experimental than under the control treatment, but the pattern of errors was similar under both conditions.

Recall. The mean numbers of items recalled on the immediate and the delayed

TABLE 3

MEAN NUMBERS CORRECT ON TESTS OF RETENTION FOR PAIRED-ASSOCIATE LISTS IN EXPERIMENT II

Condition	Categorized list		Noncategorized list	
	1.5 minutes	48 hours	1.5 minutes	48 hours
Experimental	7.8	4.3	7.7	4.5
Control	7.6	6.1	8.6	4.7

tests of retention are shown in Table 3. For the categorized material, the loss over the 48-hour period was greater under the experimental than under the control treatment. For the noncategorized material, there was a slight trend in the opposite direction. The higher order interaction, which reflects the difference in the amount of long-term proactive interference for the two types of lists, was significant ($F = 6.14$, $df = 1/120$, $p < .02$). Further analysis showed that the interaction of the condition of transfer with the length of the retention interval was significant for the categorized lists ($p < .02$) but not for the noncategorized lists.

It may be useful to restate the nature of the higher order interaction in a different way. Under the control condition, the categorized list was retained considerably better than the noncategorized list. This finding is in agreement with the results of earlier experiments which showed better retention for lists composed of mismatched normative associates rather than unrelated pairs (Postman, Fraser, & Burns, 1968; Underwood & Ekstrand, 1968). The presence of closely related items in the list appears to have a positive influence on response recall which outweighs the associative interferences at the time of the retention test. Under the experimental condition, this difference between the categorized and noncategorized materials was eliminated. According to the present interpretation, prior free recall learning increased the proactive effects of the preexisting linguistic associations sufficiently to outweigh the advantage conferred by a high level of response recall. This interpretation must be viewed with caution: the actually observed difference in long-term retention was between the control groups, and the virtual absence of a difference between the experimental groups is attributed to a balance of positive and negative factors. However, earlier analyses of extraexperimental forgetting have likewise pointed to dual and opposed effects of language habits on retention (cf. Postman, 1967).

In the recall of the categorized lists, appropriate intrusions were not numerous, but the difference that was found—12 occurrences for the experimental group and 5 for the control group—was in the expected direction. However, as in acquisition, the frequencies of such intrusions relative to all errors of commission were comparable for the two groups, namely, 17% and 14%. The relearning scores will not be reported in detail. The groups that differed on the delayed test of recall converged as relearning continued, but the initial pattern of variation remained in evidence on all but the last trial.

At this point an alternative explanation should be considered briefly. Can the higher retention of the categorized materials by control subjects be attributed to better list differentiation rather than to superior response recall? When the two lists are categorized, with no overlap of taxonomic classes between them, the successive repertoires can be distinguished more readily than when the words in each are unrelated. There are three considerations which argue against such a view. First, the same difference between unrelated pairs and scrambled associates was obtained by Underwood and Ekstrand (1968) in a study of single-list retention. In that case there was convincing internal evidence that the level of response recall was the responsible factor. Second, in the present experiment there were virtually no interlist intrusions in the recall of the unrelated words, although the number of importations from outside the list was appreciable. Third, and perhaps most important, a comparable pattern of proactive effects will come to light in the next experiment under conditions of testing which ruled out any possible influence of list differentiation.

One of the critical findings in this experiment concerns the difference between the two classes of materials. There was negative transfer in the acquisition of both types of list, but proactive interference only in the recall of categorized items. In the case of the noncategorized lists, interitem dependencies newly developed in the course of free recall learning were stable enough to impede the performance of the transfer task that followed immediately; but they apparently failed to recover after an interval of time and hence did not produce proactive losses in long-term recall. There was an alternative possibility, however, that warranted consideration: The measures of retention, which were heavily dependent on response recall, may not have been sufficiently sensitive to the proactive interferences falling on specific associations. The next study bears directly on this alternative interpretation.

EXPERIMENT III

This study again investigated the influence of prior free recall learning on paired-associate transfer and retention. As compared with the last experiment, there were two major changes. First, in order to focus as directly as possible on the associative components of transfer and interference, the test list was learned and retention was measured by the multiple-choice method. Second, only noncategorized materials were used, and a comparison was made between lists of high and of low meaningfulness. The stability of the higher order groupings established during free recall should increase as a function of the degree of meaningfulness of the words in the list. The amounts of interference in the paired-associate phase should, in turn, directly reflect the stability of the prior organization of the materials. The question of central interest was whether a test maximally sensitive to the associative component of interference would bring to light proactive effects on long-term retention even in the absence of normative relations among the items.

Method

Design. Except for the two changes noted above, the design was the same as in Experiment II. The tests of retention occurred either 3 minutes or 48 hours after the end of paired-associate learning. Thus, the design was again a $2 \times 2 \times 2$ factorial, with 2 levels of meaningfulness, 2 conditions of transfer (experimental vs. control), and 2 retention intervals.

Lists. The learning materials were four lists of 20 nouns. The words were selected from the norms of Paivio, Yuille, and Madigan (1968). The scaled meaningfulness of the words was high in two of the lists (average *m* values of 7.23 and 7.33) and low in the other two lists (average *m* values of 4.41 and 4.17). The degrees of concreteness of the words in the two types of lists were equated, but frequency of usage was correlated with *m* value. Four different lists of 10 paired associates were derived from each basic 20-word list, with the pairings again selected at random. Within each level of meaningfulness, the lists were fully balanced.

Procedure. Free recall learning in the first phase of the experiment was again to a criterion of 16/20 plus three additional study–test cycles. The words were presented on study trials at a 2.5-second rate, and the written tests of recall were 2 minutes in duration.

The paired-associate lists were learned by the multiple-choice method to a criterion of 8/10. A 2.5-second rate of presentation was used on study trials, and a 5-second rate on test trials. The interval between a study and a test trial was 5 seconds, and that between a test trial and the next study trial was 2.5 seconds. In all other respects, the multiple-choice procedure was the same as in Experiment I. The interval between the first and the second list was 5 minutes.

Retention was measured on five successive test trials. The items were again presented at a 5-second rate. The subjects received no feedback during the test of retention.

Subjects. There were 16 subjects in each of the eight groups. The source of the subjects and the manner of assignment to conditions were the same as in the previous experiments.

Results

First-list learning. In the free recall phase, the high-*m* lists were learned faster than the low-*m* lists, but the difference was not large. The mean numbers of trials to the criterion of 16/20 were 3.08 and 4.03 ($t = 4.64$, $df = 126$, $p < .01$). There were no reliable differences between groups learning the same type of list. The levels of intertrial repetition were comparable for the two types of material. For the last three tests, the mean intertrial repetition scores per trial, determined as before, were 2.93 and 2.62 for the high-*m* and the low-*m* lists, respectively.

Transfer. Acquisition of the paired-associate list by the multiple-choice method was quite rapid under all conditions (Table 4).

TABLE 4

MEAN NUMBERS OF TRIALS TO CRITERION ON TRANSFER LISTS OF HIGH AND OF LOW MEANINGFULNESS IN EXPERIMENT III

Condition	High m	Low m	$\bar{X}$
Experimental	1.94	2.41	2.18
Control	1.62	2.16	1.89
$\bar{X}$	1.78	2.28	

TABLE 5

MEAN NUMBERS CORRECT ON TESTS OF RETENTION FOR PAIRED-ASSOCIATE LISTS IN EXPERIMENT III

Condition	High m		Low m	
	3 minutes	48 hours	3 minutes	48 hours
Experimental	8.4	5.9	8.2	5.1
Control	8.9	7.7	8.9	4.2

The high-m lists were again learned reliably faster than the low-m lists ($F = 7.02$, $df = 1/120$, $p < .01$). In agreement with the results of the previous studies, there was also some evidence of negative transfer, but the difference between the experimental and the control conditions did not reach significance nor did it interact with meaningfulness. The generally rapid mastery of the task clearly limited the sensitivity of the criterion measures to differential transfer effects.[6]

Recall. The mean numbers of correct responses given on the retention test are shown in Table 5. For the high-m lists the long-term losses were, as expected, greater under the experimental than under the control treatment. The difference in the rate of forgetting was in the opposite direction for the low-m lists, indicating that the benefits of prior item familiarization outweighed whatever interference effects might have been present. Thus, the pattern of retention scores parallels that obtained in the last experiment. The three-way interaction was again significant ($F = 4.50$, $df = 1/120$, $p < .05$). When the two types of lists were considered separately, the interaction of the condition of transfer with the length of the retention interval was significant in neither case. However, in the recall of the high-m lists, with the control group having an advantage on both the immediate and the delayed test, the main effect of the condition of transfer was reliable ($p < .02$). Prior free-recall learning was, therefore, a significant source of interference, but there was no statistical support for a build-up of proactive effects over time. In view of the significance of the higher order interaction and the consistency with the results of the last experiment, the temporal trend should probably not be discounted. On the other hand, it is now clear that under the conditions of the present experiments, dependable proactive effects are obtained only when the organization of the items in the free recall phase is based on well-established linguistic dispositions.

As in the preceding experiment, there were substantial differences in long-term retention between the two control conditions. The higher the meaningfulness, the more stable were the associations formed during paired-associate learning. As a result of the proactive influence of free recall learning, this advantage was lost under the experimental conditions. This interpretation is, of course, subject to the same reservations as before. It should also be noted that in the present case, differentiation cannot be invoked as an explanatory factor, since the multiple-choice procedure entirely eliminates the requirement of list identification. On balance, the nature of the higher order interaction points once more to dual and opposed effects of prior language habits.

GENERAL DISCUSSION

The central question in this series of experiments was whether subjective groupings developed in the course of free recall learning are functionally continuous with the products of controlled associative learning. The demonstration that such groupings can become an effective source of interference in the acquisition and retention of subsequent associative tasks strongly supports an affirmative answer.

---

[6] The fact that acquisition was considerably faster than under the multiple-choice condition in the previous experiment may be attributed to the slower rate used on both presentation and test trials. A slower rate was chosen in the present experiment in order to minimize the effects of possible differences in search time for items of high and of low meaningfulness (cf. Schulz & Lovelace, 1964).

The present results do not permit us to specify in detail the mechanisms responsible for the observed interference effects. One possibility is, of course, that the subjective groupings consist of ordered sequences that come into conflict with the prescribed pairings in the transfer phase. The interference would thus reflect the successive acquisition of incompatible sequential arrangements of the items. However, there may be negative transfer from free recall to paired-associate learning—not only when the test task calls for the disruption of internally ordered units. As a result of prior experience, each item in the test list may come to arouse numerous and changing associates which enter into competition with the prescribed one. With respect to the pattern of interitem dependencies, the change would then be from a network with multiple alternative pathways to a strictly segmented one. The extent to which each of these mechanisms will come into play will, of course, depend on the nature of the organization developed in the course of free recall learning.

The finding that there was positive transfer from paired-associate to free recall learning complements the results obtained with the reverse order of tasks. The existence of stable interitem dependencies can be a source of facilitation not only when the groupings are generated by the subject but also when they are based on arbitrary pairings. Thus, even a segmented initial organization is better than none, at least in the relatively short term. It is recognized that the use of less than optimal units may under certain circumstances retard the progress of free recall learning (cf. Tulving, 1968). For purposes of the present discussion, the critical fact is that the same dispositions can function effectively in the performance of a paired-associate and a free recall task.

At this point, one may be tempted to argue that the phenomena of subjective organization can be subsumed under the principles of associative learning; or conversely, that the establishment of prescribed associations represents a special case of organization. Neither of these arguments will be entertained here. The thrust of the present comments is to call for a rapprochement between the exponents of organization and of association, with the analysis of comparable and commensurable phenomena as the meeting ground. The conceptual framework that will eventually prove most useful in the pursuit of this objective may well represent a confluence of both orientations.

Recognition of the continuity between association and organization may help us to gain a new perspective on the mechanisms of extraexperimental interference in forgetting. The basic assumption of a general theory of interference is that the principles of retroactive and proactive inhibition, as established in controlled experiments, are responsible for forgetting outside the laboratory. In the retention of verbal materials, the major sources of extraexperimental interference have been sought in the subject's language habits (Underwood & Postman, 1960). The repeated failures of predictions based on this conception have given rise to a search for the shortcomings of the theory (e.g., Postman, 1963; Underwood & Ekstrand, 1966). The possible deficiencies in the existing formulation deserve careful exploration before the hypothesis is abandoned altogether, especially in the absence of better supported alternatives. Two further questions should be added to some of the earlier speculations.

The first of these concerns the units of analysis with respect to which the subject's preexperimental dispositions are considered. We have been inclined to specify these dispositions in terms of fixed pairwise associations. Thus, the response *car* to TABLE should be difficult to learn and easy to forget because *chair* is a powerful competitor. However, pairwise linkages do not necessarily reflect verbal habits in actual use; they are usually drawn from norms gathered by the special procedures of tests of free association. Surely our associative dispositions represent highly complex networks of interword dependencies. The total pattern of interitem relations in a list is likely, therefore, to be a critical factor in acquisition and retention. Consequently, it becomes important to inquire into the conflicts that may develop between the list organization imposed in the experiment and that entailed by the subject's linguistic dispositions. Does the experimental task require the segmentation into separate and independent units of higher order groupings based on linguistic convention? The results of the experiment with categorized materials indicate that under such circumstances an appreciable amount of both associative and proactive interference may come into play.

The second question focuses on the conditions under which preexperimental dispositions may be expected to become effective in the acquisition and retention of laboratory tasks. We referred earlier to the importance of ensuring the arousal, and thus preparing the way for the subsequent recovery, of the critical linguistic dispositions in the context of the experimental situation. The results of the present studies indicate that the verbal dispositions activated in free recall learning do, in some measures, become tied to the experimental context. The requirements of organization entailed by a free recall task serve to reduce the isolation of experimental from preexperimental habits.

REFERENCES

BOUSFIELD, A. K., & BOUSFIELD, W. A. Measurement of clustering and of sequential constancies in repeated free recall. *Psychological Reports,* 1966, **19**, 935–942.

BOWER, G. H., CLARK, M. C., LESGOLD, A. M., & WINZENZ, D. Hierarchical retrieval schemes in recall of categorized word lists. *Journal of Verbal Learning and Verbal Behavior,* 1969, **8**, 323–343.

DEESE, J. Influence of inter-item associative strength upon immediate free recall. *Psychological Reports,* 1959, **5**, 305–312.

DEESE, J. Frequency of usage and number of words in free recall. *Psychological Reports,* 1960, **7**, 337–344.

EKSTRAND, B. R., & UNDERWOOD, B. J. Free learning and recall as a function of unit-sequence and letter-sequence interference. *Journal of Verbal Learning and Verbal Behavior,* 1965, **4**, 390–396.

HOVLAND, C. I., & KURTZ, K. H. Experimental studies of rote-learning theory: X. Pre-learning syllable familiarization and the length-difficulty relationship. *Journal of Experimental Psychology,* 1952, **44**, 31–39.

PAIVIO, A., YUILLE, J. C., & MADIGAN, S. A. Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology,* 1968, **76**(1, Pt. 2).

POSTMAN, L. Does interference theory predict too much forgetting? *Journal of Verbal Learning and Verbal Behavior,* 1963, **2**, 40–48.

POSTMAN, L. Short-term memory and incidental learning. In A. W. Melton (Ed.), *Categories of human learning.* New York: Academic Press, 1964.

POSTMAN, L. The effect of interitem associative strength on the acquisition and retention of serial lists. *Journal of Verbal Learning and Verbal Behavior,* 1967, **6**, 721–728.

POSTMAN, L., FRASER, J., & BURNS, S. Unit-sequence facilitation in recall. *Journal of Verbal Learning and Verbal Behavior,* 1968, **7**, 217–224.

POSTMAN, L., & STARK, K. The role of response availability in transfer and interference. *Journal of Experimental Psychology,* 1969, **79**, 168–177.

SCHULZ, R., & LOVELACE, E. A. Meaningfulness and the associative phase of paired-associate learning: A methodological consideration. *Psychonomic Science,* 1964, **1**, 37–38.

SEGAL, M. A., & MANDLER, G. Directionality and organizational processes in paired-associate learning. *Journal of Experimental Psychology,* 1967, **74**, 305–312.

SLAMECKA, N. J. An examination of trace storage in free recall. *Journal of Experimental Psychology,* 1968, **76**, 504–513.

SLAMECKA, N. J. Testing for associative storage in multitrial free recall. *Journal of Experimental Psychology,* 1969, **81**, 557–560.

TULVING, E. The effect of alphabetical subjective organization on memorizing "unrelated" words. *Canadian Journal of Psychology,* 1962, **16**, 185–191. (a)

TULVING, E. Subjective organization in free recall of "unrelated words." *Psychological Review,* 1962, **69**, 344–354. (b)

TULVING, E. Theoretical issues in free recall. In T. R. Dixon & D. L. Horton (Eds.), *Verbal behavior and general behavior theory.* Englewood Cliffs, N. J.: Prentice-Hall, 1968.

UNDERWOOD, B. J. Attributes of memory. *Psychological Review,* 1969, **76**, 559–573.

UNDERWOOD, B. J., & EKSTRAND, B. R. An analysis of some shortcomings in the interference theory of forgetting. *Psychological Review,* 1966, **73**, 540–549.

UNDERWOOD, B. J., & EKSTRAND, B. R. Linguistic associations and retention. *Journal of Verbal Learning and Verbal Behavior,* 1968, **7**, 162–171.

UNDERWOOD, B. J., & POSTMAN, L. Extra-experimental sources of interference in forgetting. *Psychological Review,* 1960, **67**, 73–95.

WOOD, G. Whole-part transfer from free recall to serial learning. *Journal of Experimental Psychology,* 1969, **79**, 540–544. (a)

WOOD, G. Whole-part transfer from paired associate to free recall learning. *Journal of Experimental Psychology,* 1969, **82**, 532–537. (b)

WOOD, G. Paired-associate and free recall to free recall transfer. *Journal of Experimental Psychology,* 1970, **83**, 519–522.

(Received September 14, 1970)

Psychological Review
1971, Vol. 78, No. 4, 303-313

FUNCTION OF DISTINCT ASSOCIATIONS FOR PAIRED-ASSOCIATE PERFORMANCE[1]

GEORGE WOLFORD[2]

Dartmouth College

A model for predicting performance on a variety of paired-associate test types is developed and empirically tested. The primary feature of the model is the assumption that subjects form distinct forward and backward associations within a single paired-associate item. Different test types have different numbers of associations available for mediating performance. Two experiments were run using recall, recognition, and a variety of forced-choice test types. Both experiments offer support for the model.

The one-element model was proposed by Bower (1961) and Estes (1961) to account for the data from a series of paired-associate experiments. The model provides a good quantitative account of the data when there are only a few well-learned responses and when learning is assessed using forward recall tests (Bower, 1961).

A number of studies which used recognition tests in addition to forward recall tests to measure learning have produced results which are not fit well by the one-element model (Estes & Da Polito, 1967; Kintsch & Morris, 1965). The two main difficulties for the one-element model are that recognition is often more superior to forward recall than can be accounted for by different probabilities of guessing correctly, and there are variables which affect recall and recognition differentially.

One possible extension of the one-element model is to assume that more than one association can be learned within a single paired-associate item. Consider the pair BOY–7. It might be possible for the subject to learn one directional association running from BOY to 7 and another directional association running from 7 to BOY. The implications of this extension of the one-element model for recall and recognition performance are developed in this paper.

DIRECTIONALITY OF ASSOCIATIONS

The learning of a single item from a paired-associate list has often been considered analogous to the learning of a classically conditioned association (Gibson, 1940; Goss, Morgan, & Golin, 1959). Such a formulation strongly implies that associations are directional when formed. In particular, an association is directed from the stimulus to the response. One major problem with the S–R formulations has been the persistent presence of backward associations. In some paired-associate experiments, the backward associations appear to equal or even exceed the forward associations in strength (Asch & Lindner, 1963; Lockhart, 1969). These strong backward associations are quite at odds with the typical difficulty experimenters in classical conditioning have had in establishing backward associations (Cason, 1935; Porter, 1938).

An interesting solution to the problem of backward associations, proposed by Asch and Ebenholtz (1962), is called the principle of associative symmetry. This principle states that upon the formation of a forward association, a backward association of equal strength is necessarily formed. Associative symmetry carries the implicit assumption that any given paired associate consists of two stimuli joined by a single bidirectional association.

---

[1] This paper is based on a PhD dissertation submitted to Stanford University (Wolford, 1969). I wish to thank my advisor, William K. Estes, for his invaluable contributions. This research was supported, in part, by Grant GB 3878 from the National Science Foundation to Stanford University and, in part, by Grant GM 16735 from the United States Public Health Service to Rockefeller University.

[2] Requests for reprints should be sent to George Wolford, Department of Psychology, Dartmouth College, Hanover, New Hampshire 03755.

One problem with the theory of associative symmetry arises from the fact that in practice backward and forward associations are rarely equal, the forward associations usually being superior, as revealed by higher recall scores (Ekstrand, 1966). To handle this difficulty, Asch and Ebenholtz (1962) further assume that both members of the pair must be of equal availability for the equality of associations to manifest itself. Availability is the ability of an item to maintain itself in short-term memory in the midst of interfering activity (Horowitz, Norman, & Day, 1966). Some experimenters have manipulated availability and obtained results in keeping with the theory of associative symmetry (Asch & Ebenholtz, 1962; Horowitz et al., 1966). Other experimenters, however, have supposedly equalized availability and the nonequality of forward and backward associations continued to appear (Coutu, 1966; Schild & Battig, 1966).

An alternative to the assumption of associative symmetry is the assumption that the subject forms distinct directional associations on any given trial. If we conceptualize a paired-associate item as consisting of two stimuli, S_1 and S_2, then two directional associations can be formed: one association running from S_1 to S_2, and another directional association running from S_2 to S_1. The function of an association is to call forth a representation of the appropriate pair member in memory. The response which is then made is dictated by the experimental situation. Therefore, instead of one bidirectional association, we would have two unidirectional associations.

A number of studies have produced evidence in favor of distinct directional associations. An experiment by Keppel and Underwood (1962) provides support for the existence of distinct forward and backward associations. They used an A–B, A–C paradigm and found that it is possible to greatly reduce the strength of forward associations without reducing the strength of backward associations. Johnston (1967) has replicated the finding with additional controls and an expanded design. If it is possible to destroy the association in one direction while leaving the association in the other direction intact, then there must have been more than one association to begin with.

Lowry and Wollen (1969) conducted an experiment in which the forward and backward associations had different retention properties. This finding provides additional support for the assumption of distinct forward and backward associations.

Experiments which find forward and backward associations to be of equal strength do not necessarily offer support for associative symmetry. There is no reason why two associations could not be distinct and still be equal in a given experiment.

IMPLICATIONS FOR PERFORMANCE

The primary reason for concern over the distinction between a single bidirectional association and two directional associations is the belief that the assumption of distinct directional associations (if valid) can increase our understanding of the interrelationships between various test types. Other experimenters have explored the implications of distinct associations in a variety of interference paradigms (Keppel, 1968; McGovern, 1964). Little attention, however, has been given to formalizing the notions or to varying test types over a wide range. First presented is an intuitive overview of the implications of the concept of distinct directional associations for a variety of paired-associate test types. Then, the formal assumptions required to make quantitative predictions about the various test types are listed.

Various types of paired-associate tests have different numbers of forward and backward associations available for mediating performance. On a forward recall test, the subject is presented with the left-hand (or temporally first) member of a pair and asked to fill in the missing right-hand member of that pair; only a learned forward association will insure correct performance. The probability of guessing correctly on a forward recall test is low or negligible in most experiments.

On an old-recognition test, the subject is presented with a complete pair and can correctly respond "old" if either the forward or backward association for that pair is learned. On a new-recognition test, the subject is presented with the left-hand pair member of one pair and the right-hand pair member of a second pair and can correctly respond "new" if the forward association of the first pair or the backward association of the second pair is learned. On an old- or new-recognition test, the subject can also guess correctly if neither association is learned. The probability of a correct guess depends on the subject's bias but is near .5 in most experiments. Recognition, then, differs from recall in potential guessing accuracy and in the potential availability of an additional association. Recognition will be most superior to forward recall when backward associations are prominent.

On one type of forced-choice test, the subject is presented with one left-hand pair member and two right-hand pair members. His task is to choose the right-hand pair member which was paired with the left-hand pair member during training. The subject has one forward and two backward associations available for mediating performance. The learned association from either member of the correct pair will insure a correct response as on an old-recognition test. In addition, if the subject has learned the backward association from the incorrect alternative, he can eliminate that alternative and respond correctly (providing all pairs within the presented list are unique). Performance on a forced-choice test with two alternatives will be superior to the average performance on old and new recognition owing to the presence of an additional association.

Two predictions have emerged thus far in the development of a model which assumes distinct directional associations. Performance on old–new recognition will be more superior to performance on forward recall tests when backward associations are strong relative to forward associations, and performance on forced-choice tests with two alternatives will be superior to the averaged performance on old–new recognition tests.

A number of previous studies have compared recognition and recall, with some studies finding recognition superior to recall (Bruce & Cofer, 1965; Postman & Rau, 1957) and some studies finding recognition equal in performance to recall, after appropriate guessing corrections (Davis, Sutherland, & Judd, 1961; Estes & Da Polito, 1967). It would be nice to show that the latter studies had less prominent backward associations, but none of the studies reported backward recall scores. Also, there are variables other than backward associations which influence comparisons between recognition and recall, such as partial associations (McNulty, 1965) and the selection of the incorrect alternatives (Bahrick & Bahrick, 1964).

The second prediction from above, namely that performance on forced choice with two alternatives will be superior to combined performance on old–new recognition, has received support in the literature (Freund, Loftus, & Atkinson, 1968; Kintsch, 1968).

OTHER THEORIES OF PAIRED-ASSOCIATE PERFORMANCE

One of the earliest theories of recognition performance held that both recall and recognition are measures of the same underlying process, but that recognition is the more sensitive of the two (Postman, Jenkins, & Postman, 1948). This view is in need of modification to account for the aforementioned cases where recall is equal to recognition. This view would be compatible with the multiple association model if "recognition is more sensitive" were translated into "recognition tests usually have more associations available for performance."

Estes and Da Polito (1967) and Kintsch and Morris (1965) have suggested that learning involves two processes: storage and retrieval. Both processes are necessary for correct performance on recall, but only storage is necessary for correct performance on recognition. It seems reasonable, however, that recognition also involves some form of retrieval. The subject must still find the association in memory in order to verify that a given pair presented for recognition is a correct one. It does not seem that the mere presence of both pair members would insure the subject's finding the association given that it is stored. It is probable, however, that the presence of both pair members will add additional cues for the finding of the association. There is also evidence in the partial information literature (McNulty, 1965) that what is stored is important for recognition-recall differences. If what is stored is only a partial association, then recognition receives more benefit than recall. In terms of the distinct association model, any pair which has been learned in the backward direction, but not in the forward direction, can be considered as a pair which is stored but is irretrievable in the forward direction.

Another relative newcomer to the area of recognition performance is signal detection theory, which was originally developed to handle the data of perceptual detection experiments and was extended to recognition memory by Egan (1958). A number of experimenters have applied the signal detection theory model to paired-associate experiments (Freund et al., 1968; Hickson, 1968; Suboski, 1967). The main problem with this theory for the present research arises in making comparisons between recognition and recall performance. For the most part, signal detection theory makes no statement about recall performance. At best it treats recall as forced choice from many alternatives. For the recall of words, two-digit numbers, or other items with large class membership, signal detection theory will generally underpredict recall when treated as forced choice. Signal detection theory and the distinct association model agree about the superiority of forced choice from two alternatives over old–new recognition.

THE MODEL

Four formal assumptions are required to incorporate the concept of distinct directional associations into a mathematical model capable of generating quantitative predictions for various paired-associate test types.

Assumption 1: The learning process for any item can be characterized as a four-state Markov process in which the four states are as follows: State U—neither association is learned; State B—only the backward association is learned; State F—only the forward association is learned; and State L—both the forward and backward associations are learned. On a single training trial, a given pair will transit from State U to State B with probability z; State U to State F with probability y; State U to State L with probability x; and the pair will remain in State U with probability $1 - x - y - z$. The transition probabilities from States B, F, and L will not enter into any of the predictions in the present paper; these transitions need to be taken into account only when there are multiple test trials with intervening training trials.

Assumption 2: Prior to the first reinforcement or training trial, all pairs are in State U.

Assumption 3: Associations within different pairs are learned independently of one another.

Assumption 4: The subject uses all available associations in making any response.

The learning process of the distinct association model (referred to as M1) described in the above assumptions can be represented by the matrix:

$$
\begin{array}{ccc}
 & & \text{State on Trial } n \\
 & & \begin{array}{cccc} \text{L} & \text{F} & \text{B} & \text{U} \end{array} \\
\begin{array}{c}\text{State}\\ \text{on}\\ \text{Trial}\\ n-1\end{array} & \begin{array}{c}\text{L}\\ \text{F}\\ \text{B}\\ \text{U}\end{array} & \begin{bmatrix} 1 & 0 & 0 & 0 \\ a & (1-a) & 0 & 0 \\ b & 0 & (1-b) & 0 \\ x & y & z & l \end{bmatrix}
\end{array}
$$

where $l = (1 - x - y - z)$.

The distinct association model is presented above in its most general form. Parameters a and b are somewhat superfluous since they do not enter into any of the predictions yet considered. They should be involved only when there are multiple reinforcement trials with intervening test trials. The correlation between the formation of the forward and backward associations is allowed to assume any value between -1 and $+1$. A correlation of $+1$ would indicate associative symmetry and would be obtained when y and z are zero. This would mean that either both associations are learned or the item remains in the unlearned state. A correlation of 0 would indicate complete independence of forward and backward associations and would be obtained when $x = (x + y) \cdot (x + z)$. This equality arises from the definition of independence, namely, two associations are independent if the probability of learning both is equal to the product of learning each separately. The actual nature of the correlation is determined by estimating the parameters from empirical data. When independence holds and the correlation between forward and backward associations is zero, the distinct association model can be rewritten as a two-parameter model.

Theoretical Predictions

Predictions for the probability correct after one reinforcement will be presented for a variety of test types.

Forward recall. The subject will be correct on a forward recall test of a given pair if that pair is in State L or F (i.e., if the forward association is learned). He can also guess correctly from State U with probability $1/r$. Therefore,

$$P(C) = x + y + (1 - x - y)1/r. \quad [1]$$

The parameter r refers to the number of members in the set from which the right-hand pair members of the list were chosen. It will be assumed in the present research that items were chosen from sufficiently large sets so that $1/r$ is approximately zero. Equation 1 can then be rewritten as:

$$P(C) = x + y. \quad [2]$$

Backward recall. It is also assumed that on backward recall tests the probability of guessing correctly from the unlearned state is approximately zero. The subject will, therefore, be correct on the backward recall of a given pair if that pair is in State L or State B.

$$P(C) = x + z. \quad [3]$$

Old recognition. On an old-recognition test, the subject is presented with both members of a pair. The subject will correctly respond "old" if the pair is in State L, State F, or State B; the subject will guess correctly from State U with probability g. Therefore,

$$P(C) = 1 - (1 - x - y - z)(1 - g). \quad [4]$$

New recognition. On a new-recognition test, the subject is presented with the left-hand member of some pair i and the right-hand member of a different pair j. He will correctly respond "new" if pair i is in State L or State F, or if pair j is in State L or State B. In other words, the subject will be correct if he has learned the forward association of pair i or the backward association of pair j. In the absence of either association, the subject will guess correctly with probability $(1 - g)$. Therefore,

$$P(C) = 1 - (1 - x - y) \times (1 - y - z)g. \quad [5]$$

Model M1 predicts different values for the probability correct on old- and new-recognition tests. Associations within different pairs (such as those available on new tests) are assumed to be learned independently of one another, but associations within the same pair (such as those available on old tests) are not necessarily learned independently of one another in Model M1.

Forward forced choice. On a forced-choice test, the subject has all of the associations which would have been available for mediating performance on an old-recognition test of the same pair plus $m - 1$ incorrect alternatives which can be eliminated on the basis of learned associations. On a forward forced-choice test (one left-hand pair member and two or more right-hand pair members), the subject eliminates the incorrect alternatives on the basis of learned backward associations. In terms of the Markov process, the subject will be correct on a forced-choice test of a pair if the pair is in State L, F, or B. If the pair is in State U, the subject will eliminate all incorrect alternatives from pairs which are in State L or B before guessing. Therefore,

$$P(C) = 1 - (1 - z - x - y) + (1 - z - x - y) \sum_{i=0}^{m-1} \binom{m-1}{i} \times (x + z)^i (1 - x - z)^{m-i-1} \frac{1}{m - i} \quad [6]$$

and II. The fit for the conditional probabilities of correct responding of Experiment I is less than adequate, but there are variables outside the realm of the model which might be contributing to the discrepancies between the predicted and observed results. These variables were mentioned above. The one-element model will not yield similarly adequate predictions of the results in Table 3. The one-element model or any similar single-process model would have to predict that if the probability of correct response on forward recall differs from one item type to another, then the probability of a correct response on recognition would also differ.

There was also some indication in the spontaneous comments and behavior of the subjects that the distinct association model may be an accurate reflection of the decision processes in these two experiments. Two subjects in Experiment II wrote in the correct associates of the incorrect alternatives on the forced-choice tests, and three subjects in Experiment I wrote in the correct associates on new-recognition tests. The writing in of incorrect alternative associations is additional evidence that subjects are using these associations as a source of information for recognition performance. Other subjects mentioned at the conclusion of the experiment that they had implicitly eliminated incorrect alternatives on new and forced-choice tests on the basis of learned associations. Mandler, Pearlstone, and Koopmans (1969) have similarly suggested that subjects may use a "retrieval check" during an old–new recognition task.

Concerning the nature of the correlation between forward and backward associations, independence seems to have the upper hand. The parameter estimates from both experiments indicate a correlation of zero between forward and backward associations. The distinct association model can be reduced to a two-parameter model in this case. Wollen, Allison, and Lowry (1969), using a different type of analysis and experiment, found a correlation somewhere between one and zero. More direct research is needed to determine the exact value of the correlation.

The distinct association model allows one to predict and vary the relationship between forward recall and recognition. Variables which strengthen backward associations relative to forward associations will result in heightened differences between forward recall and recognition. Such variables include item type (##–WORD pairs), rehearsal in the backward direction, and instructions (stressing backward recall). Converse values of the above variables will tend to lessen differences between forward recall and recognition.

The distinct association model can be extended to make predictions in any situation where multiple distinct associations are available for mediating performance. Presumably a number of interitem (and possibly intraitem) associations might be formed in other verbal learning paradigms. Intraitem associations could provide a source of information on single-item recognition tasks. Items consisting of many unique components between which such associations could be formed (such as pictures of faces) would have an advantage over sparser items. Interitem associations might be of importance in free recall. Items which have already been recalled provide cues for the recall of additional items. In order to apply the distinct association model to these situations, we need to measure the strength and determine the number of associations which are available for mediating performance in each case.

One further possibility is that changes in postcriterion latencies can be explained by the distinct association model. One learned association is sufficient for correct performance on an old-recognition test, but items which have two learned associations might have shorter latencies. Further research is clearly necessary to test the applicability of the distinct association model to the above situations.

REFERENCES

ASCH, S. E., & EBENHOLTZ, S. M. The principles of associative symmetry. *Proceedings of the*

American Philosophical Society, 1962, **106**, 135–163.

ASCH, S. E., & LINDNER, M. A. A note on "strength of association." *Journal of Psychology*, 1963, **55**, 199–209.

BAHRICK, H. P., & BAHRICK, P. O. A re-examination of the interrelations among measures of retention. *Quarterly Journal of Experimental Psychology*, 1964, **16**, 318–324.

BOWER, G. H. Application of a model to paired-associate learning. *Psychometrika*, 1961, **26**, 255–280.

BRUCE, D., & COFER, C. N. A comparison of recognition and recall in short-term memory. In, *Proceedings of 73rd Annual Convention of American Psychological Association*. Washington, D.C.: American Psychological Association, 1965.

CASON, H. Backward conditioned eyelid reactions. *Journal of Experimental Psychology*, 1935, **18**, 599–611.

COUTU, E. Associative symmetry in paired associates. *Psychonomic Science*, 1966, **5**, 389–390.

DAVIS, R., SUTHERLAND, N. S., & JUDD, B. R. Information content in recognition and recall. *Journal of Experimental Psychology*, 1961, **61**, 422–429.

EGAN, J. P. *Recognition memory and the operating characteristic*. (Tech. Note AFCRC-TN-58-51) Bloomington: Indiana University, Hearing and Communication Laboratory, 1958.

EKSTRAND, B. R. Backward associations. *Psychological Bulletin*, 1966, **65**, 50–64.

ESTES, W. K. New developments in statistical behavior theory: Differential tests of axioms for associative learning. *Psychometrika*, 1961, **26**, 73–84.

ESTES, W. K., & DAPOLITO, F. Independent variation of information storage and retrieval processes in paired-associate learning. *Journal of Experimental Psychology*, 1967, **75**, 18–26.

FREUND, R. D., LOFTUS, G. R., & ATKINSON, R. C. *Applications of multiprocess models for memory to continuous recognition tasks*. (Tech. Rep. No. 138) Palo Alto: Stanford University, Institute for Mathematical Studies in the Social Sciences, 1968.

GIBSON, E. J. A systematic application of the concepts of generalization and differentiation to verbal learning. *Psychological Review*, 1940, **47**, 196–229.

GOSS, A. E., MORGAN, C. H., & GOLIN, S. J. Paired-associates learning as a function of percentage of occurrence of response members (reinforcement). *Journal of Experimental Psychology*, 1959, **57**, 96–104.

GUTHRIE, E. R. Association as a function of time interval. *Psychological Review*, 1933, **40**, 355–367.

HICKSON, R. Signal detection in a paired-associate learning task. *Psychonomic Science*, 1968, **12**, 253–254.

HOROWITZ, L., NORMAN, S. A., & DAY, R. S. Availability and associative symmetry. *Psychological Review*, 1966, **73**, 1–15.

JOHNSTON, W. A. S–R, R–S independence and the interference potency of latent R–S associations. *Journal of Experimental Psychology*, 1967, **74**, 511–516.

KEPPEL, G. Retroactive and proactive inhibition. In T. R. Dixon & D. L. Horton (Eds.), *Verbal behavior and general theory*. Englewood Cliffs, N. J.: Prentice-Hall, 1968.

KEPPEL, G., & UNDERWOOD, G. J. Retroactive inhibition of R–S associations. *Journal of Experimental Psychology*, 1962, **64**, 400–404.

KINTSCH, W. An experimental analysis of single stimulus tests and multiple choice tests of recognition memory. *Journal of Experimental Psychology*, 1968, **76**, 1–6.

KINTSCH, W., & MORRIS, C. J. Application of a Markov model to free recall and recognition. *Journal of Experimental Psychology*, 1965, **69**, 200–206.

LOCKHART, R. S. Retrieval asymmetry in the recall of adjectives and nouns. *Journal of Experimental Psychology*, 1969, **79**, 12–17.

LOWRY, D. H., & WOLLEN, K. A. Immediate and twenty-four hour recall of S–R and R–S associations. *Journal of Experimental Psychology*, 1969, **79**, 59–62.

MCGOVERN, J. B. Extinction of associations in four transfer paradigms. *Psychological Monographs*, 1964, **78**(16, Whole No. 593).

MCNULTY, J. A. An analysis of recognition and recall processes in verbal learning. *Journal of Verbal Learning and Verbal Behavior*, 1965, **4**, 430–436.

MANDLER, G., PEARLSTONE, Z., & KOOPMANS, H. Effects of organization and semantic similarity on recall and recognition. *Journal of Verbal Learning and Verbal Behavior*, 1969, **8**, 410–423.

PORTER, J. M., JR. Backward conditioning of eyelid response. *Journal of Experimental Psychology*, 1938, **23**, 403–410.

POSTMAN, L., JENKINS, W. O., & POSTMAN, D. L. An experimental comparison of active recall and recognition. *American Journal of Psychology*, 1948, **61**, 511–519.

POSTMAN, L., & RAU, L. Retention as a function of the method of measurement. *University of California (Berkeley) Publications in Psychology*, 1957, **8**, 217–270.

SCHILD, M. E., & BATTIG, W. F. Directionality in paired-associate learning. *Journal of Verbal Learning and Verbal Behavior*, 1966, **5**, 42–49.

SUBOSKI, M. Signal detection analysis of recall paired-associates learning. *Psychonomic Science*, 1967, **7**, 357–358.

WOLFORD, G. Functions of multiple, directional associations in verbal learning. Unpublished doctoral dissertation, Stanford University, 1969.

WOLLEN, K., ALLISON, T., & LOWRY, D. Associative symmetry versus independent association. *Journal of Verbal Learning and Verbal Behavior*, 1969, **8**, 283–288.

(Received September 21, 1970)

Psychological Review
1971, Vol. 78, No. 4, 314–332

VERBAL LEARNING THEORY AND INDEPENDENT RETRIEVAL PHENOMENA[1]

EDWIN MARTIN[2]

University of Michigan

How people learn a new behavior (C) in a situation (A) for which there already exists a learned behavior (B) and how these competitive behaviors are remembered are the chief concerns of verbal learning theory. These theories entail postulated dependencies among A, B, and C. Associative interference theory has focused on A–B, A–C associative interdependence; list differentiation theory, on B, C availability interdependence; stimulus encoding theory, on stimulus identification interdependencies. The purpose of this paper is to isolate these postulated dependencies and to argue for the critical importance of finding or failing to find the requisite dependencies in memory tasks where A, B, and C are to be retrieved in some form.

The theoretic problems the verbal learners face are, in their general form, rather universal: A human learns Behavior B in Situation A. How does he do it? Later he must learn another behavior, C, which now is decreed to be appropriate to Situation A. How does he do it? At some time in the future he is confronted with Situation A. Is his behavior to be of the B or the C variety, and how so?

Answering questions like these is a theoretic enterprise: Assumptions about how people learn are arranged into arguments that might explain or predict performance in learning tasks. Critical to advancement in this enterprise is empirical evaluation of the theories proposed. Empirical evaluation in an area of science where theories are sometimes almost conversational entails two particularly important operations. One of these is clarification and definition of the collection of statements that together constitute a theory. The other is discovery of further implications of a given theory so that its acceptability can be more reliably judged.

The generality or universality of the theoretic problems addressed by verbal learners is not subverted by the scientific dictum that theories must be stated and evaluated in as cold precision as possible. Situation A and Behaviors B and C become controllably represented in the A–B, A–C transfer paradigm. The theories we seek to clarify and evaluate are stated in terms of such paradigms. The linking evaluative presumption is that any theory that proves inadequate for our simplified, controlled research situations is inadequate also for ordinary, real-life, general learning situations. Note this distinction: A theory that is inadequate for experimental results is also inadequate for general phenomena; but a theory that is adequate for experimental results may not only be inadequate for general phenomena, but irrelevant as well. This observation emphasizes the logicality of theory rejection and suggests that theory modification and refinement might best proceed in a denial instead of a confirmatory mode.

In one way or another, any theory of interest proposes dependency relations of some sort. In the transfer situation, learning to emit Response C to Stimulus A entails, by definition, learning not to emit Response B to Stimulus A. A theory that purports to explain how the learner manages such a problem must therefore postulate processes or mechanisms that interrelate B-emitting and C-emitting capabilities or probabilities. What is interest-

---

[1] Manuscript preparation was supported by the Advanced Research Projects Agency, Department of Defense (monitored by the Air Force Office of Scientific Research) under Contract AF 49(638)-1736 with the Human Performance Center, Department of Psychology, University of Michigan. The helpful comments of Judith Goggin and Arthur Melton are gratefully acknowledged.

[2] Requests for reprints should be sent to Edwin Martin, Human Performance Center, 330 Packard Road, Ann Arbor, Michigan 48104.

ing is that verbal learning theories differ in how this interdependence shall be represented. One of the purposes of this paper is to explicate as carefully as possible the locus of assumed dependency relations in two standard theories, associative interference theory and list differentiation theory. Another purpose is to draw attention to certain relevant independent retrieval phenomena to the end that these two theories might be refined or sharpened or simply discarded, either by a presently decidable verdict or through subsequent research that is sensitized to dependency assumptions (whether explicit or implicit). The final purpose of this paper is to apply this dependency approach to the rather hectically emerging stimulus encoding analyses of learning.

ASSOCIATIVE INTERFERENCE THEORY

The theoretic position of longest standing in verbal learning is associative interference theory. There are, however, several separable components of this theory, so it will be necessary to tread carefully through the various and intertwining assumptions that have come to characterize the general position. The easiest way to begin seems to be to outline the theory as it was developed and modified.

The origin of associative interference theory is traceable to G. E. Müller, who introduced the ideas of associative inhibition (Müller & Schumann, 1894) and retroactive inhibition (Müller & Pilzecker, 1900). The former idea is that learning Behavior C in Situation A (A–C learning) is more difficult after having already learned Behavior B in Situation A (A–B learning), the idea of negative transfer; the latter, that subsequent performance of Behavior B is poorer after interpolated A–C learning, the idea that A–C learning retroactively interferes with what was learned originally. Both ideas center on interference between associations that compete to produce the behavior in question, the Responses B and C.

Webb (1917) combined these two ideas (somewhat modifying the first so that it was reciprocal between A–B and A–C) into a two-factor theory of retroaction: (*a*) In transferring from original to interpolated learning in the A–B, A–C paradigm, interpolated A–C learning causes partial disorganization of what was learned in original A–B learning. (*b*) At the time of relearning A–B, the presence of Stimulus A, which was common to original and interpolated learning, serves to reinstate the interpolated as against the original association. Thus interpolated A–C learning was thought to weaken the original association between A and B and to provide the Response C as competition against whatever likelihood remained that Response B might be emitted.

McGeoch (1933a, 1933b, 1936), however, saw Webb's two factors as alternative, competing versions of a single-factor theory. He opted for the latter factor as sufficient. McGeoch's thesis was that after A–B, A–C learning, the Stimulus A can evoke both B and C but that only either B or C will be emitted according to the dominance relation between them; A–C learning does not weaken the A–B association. In his text, McGeoch (1942) summarized his position as follows:

> Inhibition is not thought of as a separate process, but as a function of competition among responses, with a resultant momentary dominance . . . of one response over another. Responses thus inhibited are not necessarily lost from the subject's repertoire, but are kept by other responses from appearing [p. 495].

In keeping with Postman's (1961, pp. 152–153) admonition, we shall label this view the *independence–dominance hypothesis* —independence of the A–B and A–C associations, with momentary dominance at the time of testing as the only cause of retroaction.

An implication of McGeoch's position is that the frequency of intrusion of the interpolated response (the C) during a subsequent test for what is remembered of original learning should correspond to the amount of observed retroaction. Melton and Irwin (1940), however, in an experiment wherein different amounts of retroaction were generated by varying the degree of interpolated learning, discovered that intrusion-error frequency and amount of retroaction did not covary as seemingly required by McGeoch. The form of discrepancy suggested unlearning of the original A–B association during interpolated A–C learning. The impact of this result was immediate, and consequently Webb's first factor was reinstated as necessary.

Underwood (1948a) observed that after having learned two associations, or rather lists of associations, in the A–B, A–C paradigm, the number of correct B responses on the first trial of A–B relearning was greater when 48 hours had elapsed between interpolated learning and relearning than when only 5 hours had elapsed. This effect he likened to spontaneous recovery as observed in conditioning experiments when a rest is introduced after extinction. Underwood (1948b) proposed that

> unlearning is to be conceived of as analogous to the extinction of a conditioned response. Like extinguished responses unlearned verbal associations recover some of their response-strength with the passage of time [p. 420].

Thus in contradistinction to McGeoch's independence–dominance hypothesis, we have the Melton-Underwood *unlearning–recovery hypothesis.*

The principal source of contention between these two hypotheses can be seen by asking the following question (which will take on more importance later): Can a given stimulus (A) have associated to it two different responses (B and C)? The independence–dominance hypothesis answers with an unqualified "yes." The pure unlearning hypothesis answers with an unqualified "no." The combined unlearning–recovery hypothesis is equivocal in that the initial answer is "no," but if the A–B association is not completely obliterated in A–C learning, it will recover over time, thereby producing a competition–dominance situation that McGeoch said holds throughout.

The unlearning–recovery hypothesis for awhile superseded the independence–dominance hypothesis in general acceptability because the latter was difficult to relate to the sometimes-observed phenomenon of spontaneous recovery. This is because the temporal course of the dominance relation was not specified. The unlearning–recovery hypothesis also was seen as relevant to the proaction phenomenon, the fact that prior A–B learning can interfere with emission of Response C after A–C learning. Since at one time it seemed that proactive effects were minimal in an immediate test after A–B, A–C learning, but increased with test delay, the recovery of unlearned A–B associations would do nicely as a workable explanation.

The modern conglomerate known as *associative interference theory* may be summarized as follows: A–B learning produces an A–B association; A–C learning produces, concomitantly, an A–C association and a weakening of the original A–B association; as time elapses, the A–B association recovers some of its strength; upon presentation of Stimulus A, the Responses B and C compete for emission in accordance with their relative associative strengths. This general view, then, is a mixture of the views of McGeoch, of Melton, and of Underwood.

The critical dependency relation in this theory is the assumption that in forming a specific A–C association, there is a concomitant weakening of the corresponding A–B association. The problem is to test this critical assumption as directly as possible.

Suppose A–B and A–C learning are paired-associate tasks where each task consists of associating pairs of verbal units. At some point in time after completion of interpolated A–C learning, the learner is presented with each A stimulus and asked to respond with both of the two responses previously associated with each stimulus. To some of the stimuli he may recall both responses (B, C); to others, only one (B, $\bar{\text{C}}$, or $\bar{\text{B}}$, C); and to some, neither ($\bar{\text{B}}$, $\bar{\text{C}}$). If learning an A–C association entails unlearning the corresponding A–B association, and if the likelihood of recalling C indexes the A–C association strength, then we must expect the following inequality: $P(\text{B}|\text{C}) < P(\text{B}|\bar{\text{C}})$; that is, it must be that recall of B is less likely when C is recalled than when C is not recalled. All of the following inequalities express this same dependency relation:

$$P(\text{C}|\text{B}) < P(\text{C})$$
$$P(\text{B}|\text{C}) < P(\text{B})$$
$$P(\text{B} \,\&\, \text{C}) < P(\text{B})P(\text{C}).$$

The last, for example, says that the likelihood of the learner recalling both B and C to Stimulus A is less than what would be expected were the A–B and A–C associations independent of or noninteractive with each other. In the event that corresponding A–B and A–C associations are in fact independent and noninteractive, all of the above inequalities would have to be rewritten as equalities to accurately describe the data.

An experiment by Abra (1969) provides a very nice set of data for testing the dependency relation that characterizes associative interference theory. Abra's subjects learned 10 pairs of verbal units in the A–B, A–C paradigm. The stimuli were consonant–vowel–consonant nonsense syllables and the responses were common two-syllable adjectives. These two lists, A–B and A–C, were each learned to a criterion of one perfect recitation. Then, either immediately or 24 hours later, the subjects were given the modified-modified free recall (MMFR) test (Barnes & Underwood, 1959). This test consists of presenting to the subject each stimulus (A) and requiring the subject to recall as best he can the two responses (B and C) previously associated with the presented stimulus. In Table 1, the mean frequencies of the four types of recall are shown (both responses; only one or the other; neither) when MMFR was immediate and when it was delayed. Thus, out of 10 stimuli for the immediate MMFR, 4.3 of them resulted in recall of both responses; 1.1, in recall of only the B response; and so on.

TABLE 1

MEAN RESPONSE FREQUENCIES

Response	Immediate MMFR			24-hour MMFR		
	C	$\overline{\text{C}}$	Total	C	$\overline{\text{C}}$	Total
B	4.3	1.1	5.4	4.2	3.5	7.7
$\overline{\text{B}}$	3.8	.8	4.6	1.4	.9	2.3
Total	8.1	1.9	10	5.6	4.4	10

Note.—Adapted from Abra, 1969.

Over the 24-hour delay, the total number of Bs recalled increased from 5.4 to 7.7, indicating a substantial amount of spontaneous recovery of the first-list responses. At the same time, the total number of Cs recalled decreased from 8.1 to 5.6. Thus the Bs and Cs appear to interact in a more-or-less pure competition–dominance manner over the MMFR delay interval. At either of the two delays, however, there is complete independence of retrieval of the Bs and Cs. For immediate MMFR,

$$P(\text{B} \,\&\, \text{C}) = .43$$

and

$$P(\text{B})P(\text{C}) = (.54)(.81) = .44;$$

for 24-hour MMFR,

$$P(\text{B} \,\&\, \text{C}) = .42$$

and

$$P(\text{B})P(\text{C}) = (.77)(.56) = .43.$$

Thus instead of an inequality, we must write, for both tests delays, $P(\text{B} \,\&\, \text{C}) = P(\text{B})P(\text{C})$.

Clearly, in Abra's experiment, delay of MMFR does not introduce dependence between B and C: B and C retrieval are no more interdependent after a 24-hour delay than in immediate MMFR. The trade off over 24 hours between Bs and Cs as revealed by the marginal totals in Table 1 (5.4 to 7.7 for the Bs, 8.1 to 5.6 for the Cs) must be attributed, therefore, to distinguishable processes that increase B availability and decrease C availability quite separately—separately, that is, in the sense that B and C do not *pairwise* interfere with the retrievability of the other.

The chi-squares for the two 2×2 subtables in Table 1 are .062 and .840, respectively.[3] With one degree of freedom,

---

[3] These are likelihood ratio chi-squares, that is,

$$\chi^2(1) = 2\sum_{i=1}^{4} O_i ln(O_i/E_i)$$

where O_i is the observed cell frequency and E_i is the cell frequency expected under the independence hypothesis. The entries in Table 1 are *mean* frequencies.

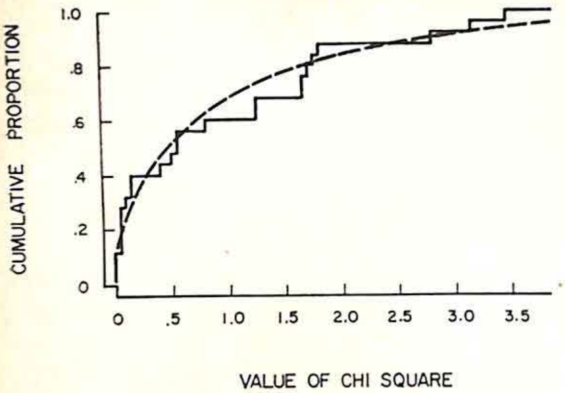


FIG. 1. Theoretical (dashed) and obtained (solid) cumulative distributions of chi-square statistics from 25 tests of response independence in MMFR tests (adapted from Greeno, 1969).

they yield $p > .80$ and $p > .30$. Thus the cell entries are completely predictable from the marginal totals. (Strictly speaking, such chi-squares may be illegitimate inasmuch as for a given 2×2 table each subject contributed to more than one cell. Thus the independent-observations assumption that underlies the chi-square test was violated to the extent that there was an interaction between subjects and items in the experiment. But such a violation can only inflate chi-square.)

This type of independent retrieval phenomenon was apparently discovered by DaPolito (1966). In three experiments, he demonstrated that competing responses were emitted independently of each other in the A–B, A–C learning paradigm. DaPolito also demonstrated that this independence of retrieval obtains whether the test is response recall (MMFR) or forced-choice recognition. Moreover, he reanalyzed the MMFR results of three standard retroaction experiments already in the literature, those of Koppenaal (1963), Postman (1964), and Postman and Stark (1965). In all of these experiments, retroaction of A–C learning against prior A–B learning was observed in subsequent MMFR tests. In the Koppenaal experiment, the MMFR test was variously delayed from the end of A–C learning for up to one week. In all instances, the independent retrieval phenomenon was observed. Thanks to the willing cooperation of other researchers, the same has been found true in the retroaction and spontaneous-recovery experiments of Birnbaum (1965), Ceraso and Henderson (1965), and Silverstein (1967) (see Martin, 1969).[4]

A summary of 25 independent tests of retrieval independence is presented in Figure 1. The dashed line is the theoretic cumulative distribution of chi-square with one degree of freedom. The solid line is the cumulative distribution of 25 observed chi-squares. This figure is from Greeno (1969), who writes:

> In all, the 25 sets of data represent quite a variety of interference experiments, with a few cases of A–Br as well as A–C paradigms, verbal stimuli as well as pictures, and words and numbers as responses. The various results also involve a wide range in the amounts of retroactive and proactive interference [pp. 17–18].

To my knowledge, there has been only one experiment designed explicitly to discover a boundary condition for this independent retrieval phenomenon. Wichawut and Martin (in press) varied the amount of A–C learning in the A–B, A–C paradigm. After learning A–B to criterion, the subjects learned A–C pairs for either 4, 8, or 12 anticipation trials; then a standard MMFR test was administered. It was discovered that the only circumstance for which $P(\mathrm{C}|\mathrm{B}) < P(\mathrm{C})$ and $P(\mathrm{B}|\mathrm{C}) < P(\mathrm{B})$ obtained was where there were only four A–C trials, and then only for those pairs where there were fewest correct anticipations in first-list learning. Otherwise, complete independence was again observed. Thus it would seem that one limitation on the independent retrieval phenomenon is that dependence appears when both B and C are barely retrievable. One might speculate that independent retrieval requires adequate representation in memory of either the A–B or the A–C pair, but not necessarily both.

Thus the dependency relation that

---

[4] Note that when recall of either the Bs or the Cs is close to perfect, the a priori likelihood of a significant chi-square is reduced. The possibility of this problem, however, does not arise in most of the many tests made; certainly not in those cases where MMFR was delayed. Moreover, were it a problem of any magnitude, the cumulative distribution of observed chi-squares shown in Figure 1 would depart from the theoretic cumulative distribution.

characterizes associative interference theory is patently denied by what we might call the DaPolito independent retrieval phenomenon: The assumption that in forming a specific A–C association there is a concomitant weakening of the corresponding A–B association is false. And given that this assumption is false, the further assumption that weakened associations recover over time does not make sense (unless it is postulated that A–B associations are weakened on some basis other than associative interference from corresponding A–C associations). In short, the independent retrieval phenomenon denies outright the unlearning–recovery hypothesis. But since there are several conversational ways to dilute this denial, some elaborative comments are in order. These comments will serve also to sharpen the claim of denial just made.

What is meant by "unlearning" in the unlearning–recovery hypothesis is a weakening of an association. If unlearning is read simply as response unavailability, then what we have is only a statement of observable fact, the fact of retroaction, not an assumption that postulates an explanative process. Demonstrations of independent retrieval have all involved relative unavailability of one or the other of the two *sets* of responses (Bs for retroaction, Cs for proaction). The independent retrieval phenomenon denies associative unlearning because the idea of associative unlearning implies a conditional *pairwise* relation between B and C availability, but the phenomenon itself is that there is no such relation. Moreover, as if to finalize this conclusion, clear changes in the relative availabilities of the two *sets* of responses (as in spontaneous recovery) do not introduce *pairwise* dependencies between responses with a common stimulus.[5]

The MMFR test was introduced by Barnes and Underwood (1959) for the express purpose of eliminating response competition (reproductive inhibition) at the time of recall: Subjects are requested to give both responses (B and C), and because presentation of the A stimuli is self-paced by the subject, there is ample time to output the nondominant response after the dominant response has been emitted. Probably the best evidence that this test is indeed free of response competition is the independent retrieval phenomenon. Pairwise associative unlearning and response competition work in the same direction, that is, both work toward the inequality $P(\mathrm{B} \,\&\, \mathrm{C}) < P(\mathrm{B})P(\mathrm{C})$. Since this inequality does not hold, both sources of its expectation can be dismissed together. Thus there can be no question of the MMFR test as possibly masking observation of the dependency relation assumed by associative interference theory.

On the basis of the preceding comments, one might overlook the relevance of McGeoch's independence–dominance hypothesis. McGeoch's hypothesis seemingly revolves around response competition and the MMFR test does not involve response competition. The hypothesis was generally applied to test situations (e.g., relearning) where only one of the two responses could be given, which means that in such situations the dominant response would be the only response emitted. In the MMFR test, both responses can be given. But even though response competition is eliminated in the MMFR, dominance remains amply evident. When MMFR is immediate, recall of the interpolated responses (the Cs) precedes, in order of output, recall of the original responses (the Bs), and as MMFR is delayed, the precedence of interpolated responses diminishes and even reverses (e.g., Koppenaal, 1963; Martin & Mackay, 1970). Thus McGeoch's independence–dominance hypothesis is certainly viable on the matter of dominance, even though McGeoch did not stipulate any general rules to govern dominance. Moreover, McGeoch called for independence between the two associational systems (original, interpolated) and has been overwhelmingly affirmed.

---

[5] Melton (1961, pp. 183–184) expressed reservations about the interpretative status of tests of the unlearning hypothesis and called for pairwise contingency analyses of the general type later carried out by DaPolito. Earlier still, Runquist (1957) demonstrated that there was no rank-order correlation between A–B pair strength and A–C pair recall, even though there was ample evidence of proaction.

In summary, the independent retrieval phenomenon is rejective of those aspects of interference theory based on common-stimulus-determined associative unlearning, but highly supportive of McGeoch's idea of associative independence. It looks as though learners do not disassociate Behavior B from Situation A when required subsequently to learn Behavior C in Situation A. We are, then, ostensibly left with the idea that indeed two different responses can be associated with a single stimulus. (The apparent counterevidence of Melton and Irwin, 1940, will be dealt with later when we consider stimulus encoding theories.) But if so, how does the learner decide when to emit B and when to emit C when Stimulus A is presented? Does the notion of response dominance help? Unfortunately, it was little more than a notion in McGeoch's hands; it does not stipulate how or when. Indeed, if dominance were not reducible to something less mindless, a subject could never learn to emit C once B were learned; and once C were learned, he would be hard pressed to emit B with increasing likelihood as time elapsed. Some mechanism is required to explain observed dominance relations.

LIST DIFFERENTIATION THEORY

That there is in fact no pairwise contingency between B and C recall in the A–B, A–C paradigm, but that there still is clear relative unavailability of Bs (retroaction) or Cs (proaction) as sets indicates that we must seek elsewhere besides specific stimulus–response associative interaction for an explanation of retroaction, proaction, and spontaneous recovery. A hypothesis of some standing that avoids the dependency defect of the unlearning–recovery hypothesis discussed above is that of list differentiation. This hypothesis also is intuitively attractive, has enjoyed a fair amount of success, and has a venerable history. In essence, it is a hypothesis that attempts to give theoretic substance to McGeoch's idea of response dominance. Since it carries a retrieval dependency implication, it falls into the purview of this paper. As before, the easiest way to explicate this hypothesis is to consider its development.

Thune and Underwood (1943) reported that after A–B, A–C learning, when the subject is presented with Stimulus A, he can often retrieve the B response but seems consciously to reject it for output purposes because he thinks it inappropriate. Underwood (1945) speculated that retroaction might therefore be at least partly a matter of response suppression on the basis of some sort of appropriateness notion on the part of the subject.

Underwood and Schulz (1960) postulated a response selector mechanism:

> The selector mechanism, itself unexplained, is assumed here as a fundamental behavioral law . . . . In the verbal-learning situation, the selector mechanism restricts the spew [output] to those response units within the list [p. 144].

In other words, A–B learning involves determination of a permissible response set, {B}. Similarly, A–C learning involves forming {C}. Negative transfer, then, is attributable to the work of differentiating Sets {B} and {C}. (Differentiation of {C} against all possible responses in general and forming A–C associations should be no more difficult than the corresponding processes in A–B learning under a pure (no associative interference) list differentiation theory.)

At the time of a test for retroaction, {B} is seen as suppressed, hence the relative unavailability of the Bs. The amount of relative unavailability of the Bs is postulated to depend on the degree of suppression of {B}. As the test for retroaction is delayed, this suppression dissipates, the availability of the Bs returns, and spontaneous recovery is observed.

Postman, Stark, and Fraser (1968) have summarized this view very plainly:

> Unlearning [retroaction] results from the operation of response selection under conditions of negative transfer and . . . this mechanism exerts its primary effect on the *entire class* of first-list responses rather than on specific stimulus–response associations [p. 689, italics added].

> This response-set interference reflects the dominance at the time of recall of the second-list criteria of selection [p. 691].[6]

---

[6] One might note that Postman refers to *selectional criteria*, an expression of the original Under-

The idea of list suppression, or how response dominance comes about, entails response-set organization in memory—the Sets {B} and {C} are memorial categories. A great deal is known about the effects on retrieval of such categorization (e.g., Cohen, 1963; Mandler, 1967). The essence of these effects is that there are definite retrieval dependencies. In a recall test, let B_n stand for recall of a B response as the *n*th output item. Then from a response-set-organization position, we must predict $P(B_n|B_{n-1}) > P(B_n)$. Where C_n is similarly defined, we must predict $P(B_n|C_{n-1}) < P(B_n)$. Yet another way to say the same thing is to write $P(B_n|C_{n-1}) < P(B_n|B_{n-1})$; that is, Bs should cluster together and Cs should cluster together.

Such inequalities hold over a wide variety of situations where experimental and normative organizations are imposed on recall tasks (ordinarily represented in terms of clustering scores of some sort). Accordingly, we have every reason to expect that after A–B, A–C learning we should observe them again were subjects asked to free recall (stimuli not present) all the Bs and Cs they can remember.

One might note that this theory has developed within the limiting confines of list-learning paradigms. Whereas associative interference theory presumed relevance for any learning situation, list differentiation theory seems fairly bound to situations wherein the to-be-learned behavior starts out as an unintegrated set of behaviors that become integrated only in the sense that nonmembers of the set are more surely excluded. The paired-associate-learning paradigm dictates that the members of this set can never become ordered, this because the cuing stimuli for the individual members appear always in random order. If one were to apply this theory to a single paired associate, he would be returned to McGeoch's position exactly, except that the word "dominant" would now read "not suppressed." When stated in this way, it is easier to see that the dependency relation that characterizes list differentiation theory is traceable to the intent of the theory to apply to list-learning paradigms.

If list organization is one way to interpret McGeoch's notion of response dominance, then so is its logical alternative. We may again consider response sets, but this time as being stimulus determined instead of list determined. Thus instead of the Sets {B} and {C}, we have the Sets $\{B_1,C_1\}$, $\{B_2,C_2\}$, $\{B_3,C_3\}$, and so on, where the subscript denotes a particular common stimulus. Under this organizational hypothesis, we must predict that in free recall of the responses, the Bs and Cs will cluster according to common stimuli, not according to list membership; that is, we must again predict retrieval dependencies, but this time on the basis of stimulus-defined response sets.

Unfortunately, very little research has been done to test these views. What little exists, however, indicates that the retrieval dependencies predicted by the list organization hypothesis do not exist. Martin and Mackay (1970) brought two lists of nine letter–word pairs to criterion in the A–B, A–C paradigm and then gave the subjects two successive free recall tests for the responses (without the stimuli present) either immediately or after a 30-minute delay. There was substantial retroaction and spontaneous recovery in the ordinary sense of unconditional availability of the B responses, but cluster analyses on the basis of list membership revealed no list organization of the responses in either the first or the second of the successive free recalls at either recall delay. Thus the retrieval dependencies predicted by the list differentiation hypothesis do not obtain; rather, we see another instance of independent retrieval, this time of the form $P(B_n|B_{n-1}) = P(B_n)$ and $P(B_n|C_{n-1}) = P(B_n)$. Had the hypothesis been true, the subject should have been more likely to remain within lists (original, interpolated) than to switch lists at any given point during the course of his recall. In fact, however, he remained within and

---

wood notion of appropriateness. To my knowledge, no theorist has postulated that subjects actually forget to which list a given response belongs; indeed, Koppenaal (1963) and others have shown that list identification accuracy falls only to 88% with as long as a one-week delay of MMFR.

switched between lists in accordance with a statistical independence hypothesis.

On the other hand, the retrieval dependencies predicted by the stimulus-determined version of the response-organization idea were variously apparent. There was significant stimulus-determined clustering in both 30-minute recalls and in the second immediate recall. Thus if $B_{i,n}$ is one of the learned responses to Stimulus A_i and is the nth output item in recall, then what was observed were the dependencies

$$P(B_{i,n}|C_{i,n-1}) > P(B_{i,n})$$

and

$$P(B_{i,n}|C_{i,n-1}) > P(B_{i,n}|C_{j,n-1}),$$

where $j \neq i$.

While it is certainly true that the results just described are not on as extensive a footing as the independent retrieval phenomenon of DaPolito, they nevertheless illustrate a definitive route toward theory refinement. If these results are reliable, then in an interference paradigm like A–B, A–C, the dominance relation between B and C cannot be a matter of list-defined interaction between sets of responses. It looks as though subjects form Sets $\{B_1,C_1\}$, $\{B_2,C_2\}$, $\{B_3,C_3\}$, and so on, instead of Sets $\{B\}$ and $\{C\}$. The implication is that learned behaviors tend to organize themselves more in terms of the stimulus features common to their past elicitation than in terms of when (e.g., in what list) they were elicited.

At first light it may seem that we have approached a dilemma: In the first section we concluded that retroaction (and proaction and spontaneous recovery) cannot be attributed to pairwise associative interference. In this section we come upon the real possibility that the Bs and Cs are organized in memory according to common stimuli, as opposed to list membership. If list membership is not to be the answer to the problems of associative interference theory, that is, if after all we find that response pairs cluster by stimuli yet do not gain their relative dominance through the associative interference that once was thought a property of the triad of two responses and one stimulus, what then? For a given stimulus A in the A–B, A–C paradigm, our apparent status is that we have B and C clustering together instead of the B with other Bs and the C with other Cs, we have a clear dominance relation between B and C (the phenomena of retroaction and proaction), and we have an absence of interference between the A–B and A–C associations.

AN IMPLICATIVE DIGRESSION

Theory in the verbal learning area has commonly dealt with three basic concepts—the stimulus, the response, and a conditional relation between them known as an association. The association has, since almost forever, been a central principle of mental organization, and associative interference theory may be considered the high-water mark of theorizing on this principle. The list differentiation theory was thus a shift in emphasis to another of the three basic concepts; it was an attempt to answer the growing difficulties associative interference theory was having with (for it) anomalous experimental results. But it too appears inadequate when one searches in data for certain predicted retrieval dependencies and cannot find them.

It is interesting, and implicative for considerations we shall take up shortly, that these two theories belong to a certain class of theories and that this class of theories is denied as a class from another quarter. They are both two-stage theories where the first stage pertains to response learning and the second, to association formation. This idea of two stages probably began with Thorndike (1932): "Other things being equal, connections are easy to form in proportion as the response is available [p. 342]" and "much of learning consists in making certain responses more available [p. 347]." Mandler (1954) carefully distinguished these two stages in an analytic fashion; and McGuire (1961) demonstrated experimentally the explanative power of Mandler's analysis. The idea that the response learning and association formation stages are temporally ordered was clearly stated by Underwood, Runquist, and Schulz (1959):

> Paired associate learning of verbal lists can be divided logically into two phases. . . . The response-learning phase is conceived of as the learning required to make the responses readily recallable. This first phase must necessarily precede the second phase, the associative phase, since the response must be available before it can be associatively connected (second phase) to a specific stimulus . . . [p. 70].

Another standard source for this hypothesis is Underwood and Schulz (1960, p. 92). From this point of view, the list differentiation hypothesis is a hypothesis that attempts to explain retroaction and spontaneous recovery in terms of the response learning (availability) stage, for it postulates the formation of Response Sets {B} and {C}, the suppression of {B} as part of the job of forming {C}, and the subsequent recovery of {B}.

An immediate implication of this view is that if somehow or other the two stages can be reliably identified, it must be that experimental variables that pertain to responses will affect the first stage, that experimental variables that pertain to stimuli will *not* affect the first stage, and that both kinds of variables will affect the second stage. Just such an argument has been made and tested by Greeno (1970). His tests were of two kinds.

The weaker test rested on the idea that the number of errors (in paired-associate acquisition) before the first correct response should index the duration of the first stage and that the number of errors after the first correct response should index the duration of the second stage. An experiment by Humphreys and Greeno (1970) provided the relevant data. They varied stimulus and response difficulty orthogonally in a between-subjects factorial design. The result was that the duration of the first stage was affected jointly (and about equally) by stimulus difficulty and response difficulty, and that the duration of the second stage was affected primarily by stimulus difficulty.

In the stronger test, Greeno fitted to the Humphreys-Greeno data a Markov model (Greeno, 1968; Humphreys & Greeno, 1970) that accurately describes a large number of detailed conditional and unconditional features of the data. One of the parameters of this model is definitely identifiable with the first stage of learning, and the effects of stimulus and response variables in the second stage can be clearly established through parameter invariance tests. The results are impressive: The first stage is jointly affected by stimulus difficulty and response difficulty, and the second stage is affected only by stimulus difficulty.

These same results were also observed in another experiment wherein different materials were used to manipulate stimulus and response difficulty (Greeno, 1970, pp. 273–275).

The preceding two sections of this paper and the results of Greeno's analysis may be summarized as follows: Not only do the associative interference and list differentiation theories predict, in their own domains, retrieval dependencies that do not exist, they are members of a particular class of two-stage theories that predict an ordered pattern of stage responsiveness to experimental variables that does not obtain in reality. Given a general commitment to the stimulus–response–association form of analysis of verbal learning, it is interesting that neither of these two theories includes a stimulus analysis and that Greeno's work indicates that one of the stages is primarily, if not totally, sensitive to stimulus variables.

STIMULUS ENCODING THEORY

Theoretic attention to the problem of how to represent the stimulus in verbal learning situations has, until recently, been sketchy at best. The general attitude has been something of the sort "a stimulus is a stimulus." Current sensitization to stimulus representation problems has probably arisen from so many different sources that a reasonable history is out of the question at this early stage. Nevertheless, some background comments will be useful in establishing the importance of this aspect of verbal learning theory, after which we can more sensibly discuss some relevant independent retrieval phenomena.

An undeniable feature of performance in a learning situation is its apparent variability. Indeed, for some theorists (e.g., Dashiell, 1937; Hull, 1930) this variability has been a critical sine qua non for learning. In general, however, the problem has been how to give this variability a workable theoretic representation. Hull's (1943, Ch. XVII) solution was to postulate a gratuitous randomly oscillating inhibitory process (${}_sO_r$) that interposed between what was stored in memory (${}_sH_r$) and the response execution apparatus. As against this rather arbitrary solution, Estes (1955a, 1955b) proposed that performance variability arises from a probabilistic relation between the learner and the stimulus situation. The idea is that over nominally identical presentations of the same stimulus event, the learner sometimes perceives one subset of component features and at other times, other subsets. Estes' original notion was that subsets were all equally likely to be sampled (or rather, all subsets of any given size were equally likely). Over the years, this sampling view was gradually modified so that by 1968 we find Trabasso and Bower arguing for selective sampling that involves progressive focusing with practice.

From another point of view, Rock (1962) has recently reminded us of, or introduced us to (as the case may be), the analyses of Höffding (1891) and Köhler (1940). Höffding conceptualized the associative relation between two external events, like S and R, as involving the associated memory traces s and r: $S \rightarrow s \rightarrow r \rightarrow R$. Höffding's message might currently be read as follows: Our theorizing has traditionally been a game played with the conditional probability $P(R|S)$. But for the subject to emit R, the trace r must be activated; for r to be activated, s must be activated; and for s to be activated, the nominal stimulus S must somehow make contact with its particular prior functional representation s. To the extent that there is any uncertainty about how nominal S will be encoded, the conditional probability $P(s|S)$ is critical; for unless the specific s to which r is associated is activated, r, and hence R, will not be forthcoming. Thus Höffding's concern, like Estes', was with variability in the perceptual relation between the learner and his learning situation.

The domain of multiple, and hence variable, encodings of a nominally unique stimulus is largely unexplored. A proper exploration, however, might provide an answer to the seeming paradox of McGeoch's well-supported contention that B and C are independently and without mutual interference associable with a common nominal stimulus in the A–B, A–C paradigm and the seeming contradiction of the Melton-Irwin and the stimulus-determined clustering results.

If XYZ–*2* and WYZ–*9* are two paired associates to be learned in the same experimental task, it is obvious that a rational subject will associate *2* with X and *9* with W. In an important and germinal paper, Underwood (1963) pointed out that learners will sensibly disregard those components of composite stimuli that are nonessential to the task at hand. Postman and Greenbloom (1967) and several others have demonstrated that, in essence, subjects will seldom make use of all the distinguishing features of stimuli if criterional performance does not require it. These are instances in which subsequent experimental tests have shown that learners isolate particular and definable components of a nominal stimulus for use as cues for adaptive overt behavior. It is for such circumstances as these that I shall shortly cite some implicative retrieval dependency data.

In distinction to these explicit selectional results, there is the case where a given nominal stimulus is perceived accurately and in its entirety, yet is internally or implicitly or cognitively interpreted in only one of two or more clearly disjoint ways. In this important case, we know nothing about retrieval dependence or independence. Suppose, as an example, that the auditory signal /stil/ is presented to a subject as the nominal stimulus S. Let s_i be the resultant interpretation by the subject, where s_1 is *steel* and s_2 is *steal*. From the work of Galbraith and Taschman (1969), we know normatively that $P(s_1|S) = .37$ and $P(s_2|S) = .63$. A hypothesis that grants to s_1 and s_2 an elicitative, associative relation must predict that subsequent visual recognition of STEEL will positively covary with recognition of STEAL. A hypothesis that s_1 and s_2 are mutually inhibitory must predict that recognition of STEEL will negatively covary with recognition of STEAL. An independence hypothesis predicts that once *steel* and *steal* are entries in immediate memory, their retrieval is independent. Such dependencies are testable and must certainly be implicative for theories about how semantic and associative spaces are organized.

Also within the general category of implicit stimulus encoding is the intuitively satisfactory result first reported by Underwood (1965). The nominal stimulus TABLE is associatively elaborated to include *chair* so that when later CHAIR is presented in a recognition test, it is falsely recognized as having been previously presented. But the question that has not been addressed is whether recognition of TABLE and CHAIR is stochastically dependent. Is *chair* memorially substituted for *table* so that recognition for TABLE is dependently suppressed? Is recognition for TABLE enhanced because *chair* was associatively addended to *table*? Or is it the case that once *chair* is evoked as a memorial image, *table* and *chair* are independent entities in immediate memory? Inasmuch as this paper is about retrieval dependencies or the absence of such dependencies vis-à-vis verbal learning theories, this seems the place to insist that their delineation is critical, not only to a proper evaluation of current theory but also to a satisfactory approach to new theory.

Let us tentatively assume that stimulus selection phenomena (e.g., Postman & Greenbloom, 1967; Underwood, 1963) refer to more or less external, attentional responses to particular components of composite stimuli in a paired-associate task. The effectiveness of each of the components as an elicitor of the response has been determined in two ways, the subsequent power of an isolated component to elicit the response (e.g., Postman & Greenbloom, 1967) and the subsequent differential power of the response to elicit backwardly some of the components (e.g., Hintzman, 1969). While there is thus a clear case for stimulus component selection, no theoretic advance has been made if it turns out that the components selected by the subject are themselves mutually associated and bound together; for if they are, then the only result is that the observer (the experimenter, the theorist) was mistaken as to what constituted a whole and indivisible stimulus. On the other hand, if it turns out that the individual components selected are *not* bound together, either associatively or in some Gestaltish way, then the problems raised in this paper become more tractable.

Wichawut and Martin (1970) and Martin[7] have demonstrated that stimulus components do not interassociate with each other. Wichawut began with two alternative hypotheses, namely, that if $A_1A_2A_3$ is the composite nominal stimulus, then either A_i becomes independently associated with Response B or the subjectively selected A_is become associated with each other and the integrated package becomes associated with B as a whole.

In Wichawut's experiment, each A_i and B was a four-letter noun, and three different groups of 32 subjects each were given three widely spread degrees of A–B learning so they could variously do what they might about stimulus encoding and organization. (The lowest degree of learning was just meeting an 8/8 criterion, that is, a single perfect recitation of the eight-pair list. The other degrees of learning entailed either 8 or 16 additional learning trials.) After A–B learning, where $A = \{A_1A_2A_3\}$, the subjects were presented with either A_1, A_2, A_3, or B from each of the A–B pairs and asked to provide the missing three members.

If it is true that stimulus components become associated with each other during learning, then Component A_i should elicit Component A_j, $j \neq i$, with some probability. But there are two ways A_i could elicit A_j, either directly via an A_i–A_j association, or indirectly via mediation through the Response B, A_i–B–A_j. Accordingly, when A_i is the presented stimulus component, there are two cases to consider: when the subject recalls Response B and when he

---

[7] E. Martin. Stimulus component independence. In preparation.

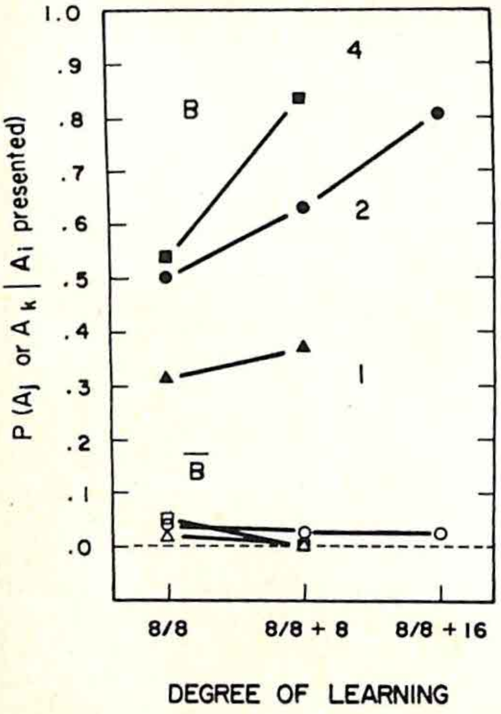


FIG. 2. Proportion of times when at least one stimulus component (A_j or A_k) other than the one presented (A_i) was recalled given recall (upper curves, B) and nonrecall (lower curves, $\bar{B}$) of the response, as a function of degree of learning. (Parameter is length of anticipation interval in seconds.)

fails to recall Response B. We assume that if he cannot recall B, then he cannot recall A_j through the mediative chain A_i–B–A_j.

In Figure 2, the upper curve with filled circles (the one with the number 2 beside it) shows the proportion of times when in addition to B being recalled, at least one stimulus component (A_j, A_k) other than the one presented (A_i) was also recalled. As is evident, when Response B was recalled, recall of other stimulus components increased with degree of A–B learning. On the other hand, the lower curve with open circles shows the proportion of times when at least one stimulus component other than the presented component was recalled when Response B was *not* recalled. As this curve indicates, Stimulus Component A_i in no way suggested its fellow components, A_j and A_k, in the case where Response B could not be recalled. Stimulus components apparently do not associate directly with each other; the retrievability of A_j when A_i is presented must be mediated through their common response, B.

Wichawut's experiment involved an exposure of two seconds of the compound stimulus in an anticipation learning procedure. Mackay varied this exposure, allowing either one second or four seconds to study or do with each stimulus whatever subjects do with stimuli. He imposed two degrees of A–B learning, the 8/8 criterion and the 8/8-plus-8 conditions of Wichawut. His results are shown in Figure 2 by filled triangles for one-second and filled squares for four-second stimulus study times for the case where the Response B was recalled (upper curves), and by open triangles and squares, respectively, when the Response B was not recalled (lower curves). As can be seen from the upper curves, the amount of stimulus study time (1, 2, 4 seconds) clearly affected the indirect, response-mediated associations between stimulus components. Equally clear, however, is the fact shown by the lower curves: In the absence of response recall, there is no retrieval of other stimulus components.

The subjects in these experiments knew nothing of the later components tests. And in perfect keeping with standard results, they differentially utilized the three components of the stimuli during acquisition (see the articles). Yet when their behavior was analyzed for stimulus intercomponent dependencies, none were to be found. Irrespective of degree of learning, these subjects had no way of getting from one stimulus component to another except via the assigned response that was common to all the components. The open-figure data in Figure 2 are so consistently close to zero that the idea of a mistake is probably unreasonable. If anything, independence among stimulus components increased with degree of learning, an observation that is reminiscent of the boundary condition for the DaPolito independent retrieval phenomenon noted in the first section of this paper.

These results require generalization; but their initial implication is clear: Stimulus components that are a priori independent do not become interdependent or associated through learning, at least not in a rote experimental setting. Mediation through a common response can give the appearance of interdependence, but a proper analysis shows that after all it is not so.

In the spirit of the inchoate nature of stimulus encoding theory, only a few evaluative comments are in order. For one, Estes' (1955a, 1955b) general idea that stimulus elements are, at some level of analysis, independently processed into memory seems correct. For another, Underwood (1963) and Trabasso and Bower (1968) are clearly correct that selective (focus) sampling is what happens. (Estes' component-independence idea and the facts of selective sampling are not mutually contradictory as long as a given stimulus component is selected on some basis other than the subject having already selected or rejected some other component.) And certainly Höffding (1891) was justified in his fears that stimulus identification is a risky affair (see Martin, 1968, for an elaboration of this point). While these conclusions hardly constitute a definitive picture of how stimuli get represented in memory, they are suggestive of a way out of some of the difficulties in verbal learning theory.

INTEGRATION

We noted at the outset three general questions: What is learned when a learner learns Behavior B in Situation A? How does some new, different Behavior C displace Behavior B in Situation A? And given Situation A at some later point in time, what will be remembered of Behaviors B and C? Probably one of these questions is more important than the other two, and it would appear that it is the second question, the question of transfer, that is critical. In fact, it is with respect to the juncture of A–B and A–C learning that the two theories discussed earlier generate their retrieval dependency statements for what should happen in a later test for what is remembered of A–B and A–C learning. Moreover, an answer to this question must take cognizance of some learning process that is central to both A–B and A–C learning.

An opening answer to this question is that the A in A–C learning is not the same A as in A–B learning. For the purpose of discussion, let A_B and A_C be the subjective, functional versions of nominal A in A–B and A–C learning, respectively. The critical problem is to specify how A_C differs from A_B and how the difference comes about in learning.

Begin with the eminently sensible notion that Nominal Stimulus A is multifeatured (e.g., Anisfeld & Knapp, 1968; Bower, 1967; Underwood, 1969; Wickens, 1970). Although stimulus selection research indicates that obvious, externally definable components comprise a share of the features in some experimental situations, they are by no means exhaustive of the list. Many other properties of a stimulus are also important, as frequency and temporal point of occurrence (Underwood, 1969) as well as their modality of occurrence, class membership, and so on (Wickens, 1970). For a given Nominal Stimulus A, let A be described by the features $a_1, a_2, \cdots, a_n$. Moreover, assume that these features are differentially salient, that is, not equally likely to be sampled (noticed, registered) by the learner.

The features a_i of nominal A may be viewed as coordinates of a vector. Over repetitions within the same task, as in A–B learning, some of these coordinates will change in value, others will remain constant. For example, one of the a_i may be a frequency counter and hence change with trials, another a_i may be an index of recency (Underwood, 1969). The a_i that describe the phonetic and graphemic natures of A probably will not change over trials. Given both that there can be variation in which features are sampled and that some features (coordinates) change value over trials, it follows that the functional representation of nominal A is different even between the beginning and end of A–B learning.

But in the A–B, A–C paradigm, nominal A carries on throughout, and hence there may seem no way to make A_C different from A_B in any way sufficiently distinctive to account for transfer, retroaction, proaction, and spontaneous recovery. Two interrelated proposals have recently been made. The author (Martin, 1968) has suggested that the sampling probability distribution over the feature set $\{a_i\}$ changes when A–C learning is encountered after A–B learning; that if a_i is the most salient feature in A–B learning, then somehow a_j becomes the most salient feature in A–C learning. Learning was seen as associating the currently prescribed response to the most salient feature. Transfer was explicated in terms of the numerosity of alternative features: The more features descriptive of the stimulus, the more different (in terms of degrees of freedom) can the A–C distribution be from the A–B distribution. On this view, retroaction is a subjective persistence of the feature sampling bias of A–C learning.[8] Proaction is accountable to there remaining a residual likelihood of sampling the feature that was most dominant in original A–B learning. As the test (e.g., MMFR) is delayed from the end of A–C learning, the feature-sampling distribution as biased by A–C learning was presumed to revert to its original (preexperimental?) shape, thus reducing the dominating likelihood of Response C over Response B (spontaneous recovery).

The soft spot in this theory is flagged by the word "somehow" in the phrase "somehow a_j becomes the most salient feature in A–C learning." The second proposal alluded to above suggests a way to differentiate A_C from A_B. Greeno (1970) has proposed that the first stage of acquisition is storing in memory the stimulus–response pair itself, as a unit (as opposed to storing just the R). This proposal is immediately attractive for two reasons: First, it makes sense in view of the fact discussed earlier that the first stage of learning is affected by both stimulus and response variables (the source of Greeno's idea to start with). Second, by placing the stimulus and the response together as adjacent, interacting components of a single entry into memory, it makes the response a highly likely determiner of which stimulus features will be attended, selected, or emphasized. And because there is no greater difference between A–B and A–C learning than the difference between B and C, B and C themselves could be the primary source for the difference between A_B and A_C.[9] If this is how things work, then A_B and A_C should be more surely distinguished to the extent that B and C belong to distinct classes of items, with the result that less retroaction will occur. Happily, this is true in fact (Birnbaum, 1968; Friedman & Reynolds, 1967; Postman, Keppel, & Stark, 1965; Shulman & Martin, 1970).

The second stage in a verbal learning task is establishing a retrieval route from the Nominal Stimulus S to the unitary memory entry, the stimulus–response pair (Greeno, 1970). In the A–B, A–C interference paradigm, the Responses B and C contribute to the distinguishability of these retrieval routes, this through their influence on emphases of the features of nominal A. Given that B and C are responsible (at least in part) for the difference between A_B and A_C, how can we explain the DaPolito independent retrieval phenomenon, the fact that B and C cluster together instead of with other Bs and Cs, and Melton's seeming counterevidence against McGeoch's independence hypothesis?

Begin by transforming the Wichawut stimulus component independence phenomenon into a premise that applies to stimulus features in general. To do so is perfectly reasonable. For example, the graphemic feature G of the stimulus GEORGE can hardly be associated with the semantic animacy feature of GEORGE, any more than can the phonetic signal /in/ be selectively associated with the semantic encoding of INTER or SIN.[10] Assume that upon receipt of a stimulus, the subject draws a random sample of the features that describe that stimulus such that the paired response is a determiner of the probability distribution over those features. Then three points can be argued:

---

[8] This constitutes an interpretation of the generalized response competition idea of Newton and Wickens (1956), which holds that subjects persist in their most recent behavior. Inasmuch as stimuli are the excuse for behavior, persistence is here seen as a matter of how the stimuli are encoded.

[9] A convincing example is the professor who may in a given meeting of his Senate act either as a substantive participant or as the parliamentarian. Clearly, the behavior expected of him will determine which features of the proceedings he selectively attends to and keeps track of.

[10] Beginning with Miller and Nicely (1955) and Liberman (1957), there has been a general under-

1. If A_{B_1} and A_{B_2} are two stimuli in A–B learning and A_{C_1} is the A–C learning version of A_{B_1}, then certainly A_{B_1} will be more similar to A_{C_1} than to A_{B_2} (in standard paired-associate experiments where intralist similarity is minimized). Accordingly, if in free recall (stimuli absent) of the B and C responses the Response B_1 is recalled, then via mediation through A_{B_1} the subject is more likely to follow up with recall of C_1 than with B_2. Hence the stimulus-determined clustering results of Martin and Mackay (1970). (Of course, the pictures for when intralist stimulus similarity is high and when the paradigm is A–B, C–D must be appropriately different.)

2. To simplify matters for the sake of argument, suppose Nominal Stimulus A has features a_1, a_2, a_3, and a_4 only. Let $A_B = a_1$ and $A_C = a_2$ as a result of selective stimulus encoding in A–B and A–C learning, respectively. In the MMFR test, suppose the subject takes a size-two random sample of these features. If this sample includes both a_1 and a_2, he will recall both B and C; if a_1 and either a_3 or a_4, he will recall only B; if a_2 and either a_3 or a_4, then only C; if a_3 and a_4, then neither B nor C. Since this sample is random, there will be no retrieval dependency between Responses B and C. Hence the DaPolito independent retrieval phenomenon. The apparent dominance of C over B in retroaction is dictated solely by the shape of the probability distribution over $\{a_i\}$. The amount of proaction follows similarly, from the likelihood that (in this example) a_1 but not a_2 will be sampled.

3. The generalized Melton-Irwin (1940) phenomenon is that overt intrusion of the interpolated Cs in a subsequent test for what is remembered of original A–B learning first increases and then decreases with degree of interpolated A–C learning, while retroaction increases throughout.[11] In light of the point of view being argued, the explanation would be that with limited A–C practice, the subject is still, on occasion, retrieving the original A–B unit from memory—this as opposed to unlearning the association between A and B. As A–C trials accumulate, A_C becomes differentiated from A_B so that the nominal A is interpreted reliably as A_C, to the end that only the unit A–C is retrieved from memory. The everincreasing degree of retroaction is simply the increasing stimulus encoding bias toward A_C instead of A_B.

What the dependency relation that characterizes this view should be is not clear at this stage of development. With some confidence we can say that it is not associative dependency among stimulus components. The central postulate, however, is that A_C is different from A_B and that A–C learning involves differentiation of A_B and A_C. The idea that A_C might differ from A_B has received some attention in the literature. The results have been mixed, but they are worth commenting on. After having learned a list of A–B pairs, Houston's (1967, Experiment III) subjects were then required to learn a second list of the form either AX–C or AX–C′, where X was a new stimulus component (added to the original A) and C and C′ were new responses that were, respectively, different or meaningfully similar to the original B responses. Following this second task, his subjects were presented with either A, X, or AX and requested to provide the second-task response (C or C′). For the AX–C′ condition of second-list learning, the compound AX stimulus elicited 7.00 (out of a possible 8) C′ responses. The X and A components elicited, respectively, 2.88 (41% of 7.00) and 5.88 (84%) C′ responses, indicating that when the first- and second-task responses were similar (one possibly

---

standing that speech perception depends on combinations of perceptually independent features (probably articulatory, see Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967).

[11] The Melton-Irwin result was observed not in paired-associate but in serial learning. The basic reference for this phenomenon in paired-associate learning should be Underwood (1945).

mediating the other), subjects tended to stick with the first-task stimulus (the A component). But when the first- and second-task responses were quite dissimilar (the AX–C condition), the respective mean number of second-task responses elicited were 7.13 for AX, 4.38 (61%) for X, and 4.88 (68%) for A, indicating that in an interference situation there was a tendency to switch to the new component (X) or to add the new component to the old component (A). The critical comparison is for the X component: 41% for the AX–C′ condition, 61% for the AX–C condition. An ordinary 2 × 3 analysis of variance of these data did not, however, yield a statistically significant interaction, the fact that led Houston to conclude against recoding of the stimulus in the second task.

But results from such experiments as Houston's, whatever their apparent outcome, must be viewed with reservation. Implicit in the conclusions drawn from them is the idea that the functional encoding of a given stimulus is to be identified with, say, a particular letter of that stimulus. Surely this is an idea we can no longer entertain seriously, especially in light of the numerous encoding possibilities already listed with firm, supporting evidence (e.g., Underwood, 1969; Wickens, 1970). As if to emphasize the difficulty in trying to identify obvious, external, experimenter-defined stimulus attributes with the learner's subtle, internal, subjectively defined functional encodings, we may note that Schneider and Houston (1968) found positive evidence for shifts in stimulus encoding in a negative transfer paradigm while Williams and Underwood (1970, Experiment I) did not.

The general idea of stimulus recoding on the part of a learner who is faced with learning a new behavior in an old situation must be distinguished from an accompanying and complementary notion, namely, that one of the products of first-task learning in the experimental paired-associate list situation is reliable differentiation among the several stimuli in the list. Presumably, this accomplishment requires effort; and if so, learners may well prefer to make use of this differentiation in the second task, either by exercising some variation on the original encoding or combining the original encoding with some new feature. If this is the case, then rough tests of recoding not only can fail to provide evidence that recoding occurred but also can provide, spuriously, evidence that the original encoding was retained in its original form in second-task learning. In short, it seems to be true that we have, on occasion, allowed conveniences of method and measurement to play the role of silent or hidden assumptions in our inferences.

CONCLUSIONS

The general view outlined in the preceding section is in agreement with McGeoch's hypothesis of independence between A–B and A–C associations. It proposes that McGeoch's idea of dominance be interpreted as a stimulus encoding bias that is in part determined by the to-be-associated response. It says "yes" that two distinct behaviors can be associated to the same nominal stimulus situation, but "no" that they can be associated to the same functional encoding of that situation.

The general view argued in the preceding section is that a major portion of learning is perceptual learning—learning an effective identifying response to the nominal stimulus situation. Learning new behavior in the same nominal situation (transfer) is, accordingly, largely a matter of learning to perceive the situation differently from before. The phenomena of retroaction, proaction, and spontaneous recovery are thus seen as determined by how the learner perceives (encodes) the stimulus situation on a later test occasion.

In general perspective, the positions taken in this paper are in part intended to shift the focus of theoretic attention to the perceptual end of the human information-processing system, a seemingly more fertile domain; a domain, moreover, in greater heuristic contact with other areas of psychology. Of course, it does not follow, pari passu, that old problems are thereby solved; only that there may be a more likely route of approach.

REFERENCES

Abra, J. C. List-1 unlearning and recovery as a function of the point of interpolated learning. *Journal of Verbal Learning and Verbal Behavior*, 1969, **8**, 494–500.

Anisfeld, M., & Knapp, M. Association, synonymity, and directionality in false recognition. *Journal of Experimental Psychology*, 1968, **77**, 171–179.

Barnes, J. M., & Underwood, B. J. "Fate" of first-list associations in transfer theory. *Journal of Experimental Psychology*, 1959, **58**, 97–105.

Birnbaum, I. M. Long-term retention of first-list associations in the A–B, A–C paradigm. *Journal of Verbal Learning and Verbal Behavior*, 1965, **4**, 515–520.

Birnbaum, I. M. Response-class similarity and first-list recall with mixed and unmixed transfer designs. *Journal of Experimental Psychology*, 1968, **77**, 542–546.

Bower, G. A multicomponent theory of the memory trace. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation*. Vol. 1. New York: Academic Press, 1967.

Ceraso, J., & Henderson, A. Unavailability and associative loss in RI and PI. *Journal of Experimental Psychology*, 1965, **70**, 300–303.

Cohen, B. H. Recall of categorized word lists. *Journal of Experimental Psychology*, 1963, **66**, 227–234.

DaPolito, F. J. Proactive effects with independent retrieval of competing responses. Unpublished doctoral dissertation, Indiana University, 1966.

Dashiell, J. F. *Fundamentals of general psychology*. New York: Houghton, 1937.

Estes, W. K. Statistical theory of spontaneous recovery and repression. *Psychological Review*, 1955, **62**, 145–154. (a)

Estes, W. K. Statistical theory of distributional phenomena in learning. *Psychological Review*, 1955, **62**, 369–377. (b)

Friedman, M. J., & Reynolds, J. H. Retroactive inhibition as a function of response-class similarity. *Journal of Experimental Psychology*, 1967, **74**, 351–355.

Galbraith, G. G., & Taschman, C. S. Homophone units: A normative and methodological investigation of the strength of component elements. *Journal of Verbal Learning and Verbal Behavior*, 1969, **8**, 737–744.

Greeno, J. G. Identifiability and statistical properties of two-stage learning with no success in the initial stage. *Psychometrika*, 1968, **33**, 173–215.

Greeno, J. G. *A cognitive interpretation of negative transfer and forgetting of paired-associates*. (Human Performance Center Memorandum Rep. No. 9) Ann Arbor: University of Michigan, 1969.

Greeno, J. G. How associations are memorized. In D. A. Norman (Ed.), *Models of human memory*. New York: Academic Press, 1970.

Hintzman, D. L. Backward recall as a function of stimulus similarity. *Journal of Verbal Learning and Verbal Behavior*, 1969, **8**, 384–387.

Höffding, H. *Outlines of psychology*. (Trans. by M. E. Lowndes) London: Macmillan, 1891.

Houston, J. P. Stimulus selection as influenced by degrees of learning, attention, prior associations, and experience with the stimulus components. *Journal of Experimental Psychology*, 1967, **73**, 509–516.

Hull, C. L. Simple trial-and-error learning: A study in psychological theory. *Psychological Review*, 1930, **37**, 241–256.

Hull, C. L. *Principles of behavior*. New York: Appleton-Century-Crofts, 1943.

Humphreys, M., & Greeno, J. G. Interpretation of the two-stage analysis of paired-associate memorizing. *Journal of Mathematical Psychology*, 1970, **7**, 275–292.

Köhler, W. *Dynamics in psychology*. New York: Liveright, 1940.

Koppenaal, R. J. Time changes in the strengths of A–B, A–C lists: Spontaneous recovery? *Journal of Verbal Learning and Verbal Behavior*, 1963, **2**, 310–319.

Liberman, A. M. Some results of research on speech perception. *Journal of the Acoustical Society of America*, 1957, **29**, 117–123.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. Perception of the speech code. *Psychological Review*, 1967, **74**, 431–461.

Mandler, G. Response factors in human learning. *Psychological Review*, 1954, **61**, 235–244.

Mandler, G. Organization and memory. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation*. Vol. 1. New York: Academic Press, 1967.

Martin, E. Stimulus meaningfulness and paired-associate transfer: An encoding variability hypothesis. *Psychological Review*, 1968, **75**, 421–441.

Martin, E. *Associative interference theory and spontaneous recovery*. (Human Performance Center Memorandum Rep. No. 10) Ann Arbor: University of Michigan, 1969.

Martin, E., & Mackay, S. A. A test of the list-differentiation hypothesis. *American Journal of Psychology*, 1970, **83**, 311–321.

McGeoch, J. A. Studies in retroactive inhibition: I. The temporal course of the inhibitory effects of interpolated learning. *Journal of General Psychology*, 1933, **9**, 24–43. (a)

McGeoch, J. A. Studies in retroactive inhibition: II. Relationships between temporal point of interpolation, length of interval, and amount of retroactive inhibition. *Journal of General Psychology*, 1933, **9**, 44–57. (b)

McGeoch, J. A. Studies in retroactive inhibition: VIII. Retroactive inhibition as a function of the length and frequency of presentation of the interpolated lists. *Journal of Experimental Psychology*, 1936, **19**, 674–693.

McGeoch, J. A. *The psychology of human learning*. New York: Longmans, 1942.

McGuire, W. J. A multiprocess model for paired-associate learning. *Journal of Experimental Psychology*, 1961, **62**, 335–347.

Melton, A. W. Comments on Professor Postman's paper. In C. N. Cofer (Ed.), *Verbal learning and verbal behavior.* New York: McGraw-Hill, 1961.

Melton, A. W., & Irwin, J. McQ. The influence of degree of interpolated learning on retroactive inhibition and the overt transfer of specific responses. *American Journal of Psychology*, 1940, **53**, 173–203.

Miller, G. A., & Nicely, P. E. An analysis of perceptual confusions among some English consonants. *Journal of the Acoustical Society of America*, 1955, **27**, 338–352.

Müller, G. E., & Pilzecker, A. Experimentelle Beiträge zur Lehre vom Gedächtnisses. *Zeitschrift für Psychologie*, Ergänzungsband 1, 1900.

Müller, G. E., & Schumann, F. Experimentelle Berträge zur Untersuchung des Gedächtnisses. *Zertschrift für Psychologie*, 1894, **6**, 81–190, 257–339.

Newton, J. M., & Wickens, D. D. Retroactive inhibition as a function of the temporal position of interpolated learning. *Journal of Experimental Learning*, 1956, **51**, 149–154.

Postman, L. The present status of interference theory. In C. N. Cofer (Ed.), *Verbal learning and verbal behavior.* New York: McGraw-Hill, 1961.

Postman, L. Studies of learning to learn: II. Changes in transfer as a function of practice. *Journal of Verbal Learning and Verbal Behavior*, 1964, **3**, 437–447.

Postman, L., & Greenbloom, R. Conditions of cue selection in the acquisition of paired-associate lists. *Journal of Experimental Psychology*, 1967, **73**, 91–100.

Postman, L., Keppel, G., & Stark, K. Unlearning as a function of the relationship between successive response classes. *Journal of Experimental Psychology*, 1965, **69**, 111–118.

Postman, L., & Stark, K. The role of response set in tests of unlearning. *Journal of Verbal Learning and Verbal Behavior*, 1965, **4**, 315–322.

Postman, L., Stark, K., & Fraser, J. Temporal changes in interference. *Journal of Verbal Learning and Verbal Behavior*, 1968, **7**, 672–694.

Rock, I. A neglected aspect of the problem of recall: The Höffding function. In J. M. Scher (Ed.), *Theories of the mind.* New York: Free Press, 1962.

Runquist, W. N. Retention of verbal associates as a function of strength. *Journal of Experimental Psychology*, 1957, **54**, 369–375.

Schneider, N. G., & Houston, J. P. Stimulus selection and retroactive inhibition. *Journal of Experimental Psychology*, 1968, **77**, 166–167.

Shulman, H. G., & Martin, E. Effects of response-set similarity on unlearning and spontaneous recovery. *Journal of Experimental Psychology*, 1970, **86**, 230–235.

Silverstein, A. Unlearning, spontaneous recovery, and the partial-reinforcement effect in paired-associate learning. *Journal of Experimental Psychology*, 1967, **73**, 15–21.

Thorndike, E. L. *The fundamentals of learning.* New York: Columbia University Press, 1932.

Thune, L. E., & Underwood, B. J. Retroactive inhibition as a function of degree of interpolated learning. *Journal of Experimental Psychology*, 1943, **32**, 185–200.

Trabasso, T., & Bower, G. H. *Attention and learning.* New York: Wiley, 1968.

Underwood, B. J. The effect of successive interpolations on retroactive and proactive inhibition. *Psychological Monographs*, 1945, **59**(3, Whole No. 273).

Underwood, B. J. Retroactive and proactive inhibition after 5 and 48 hours. *Journal of Experimental Psychology*, 1948, **38**, 29–38. (a)

Underwood, B. J. "Spontaneous" recovery of verbal associations. *Journal of Experimental Psychology*, 1948, **38**, 429–439. (b)

Underwood, B. J. Stimulus selection in verbal learning. In C. N. Cofer & B. S. Musgrave (Eds.), *Verbal behavior and learning.* New York: McGraw-Hill, 1963.

Underwood, B. J. False recognition produced by implicit verbal responses. *Journal of Experimental Psychology*, 1965, **70**, 122–129.

Underwood, B. J. Attributes of memory. *Psychological Review*, 1969, **76**, 559–573.

Underwood, B. J., Runquist, W. N., & Schulz, R. W. Response learning in paired-associate lists as a function of intralist similarity. *Journal of Experimental Psychology*, 1959, **58**, 70–78.

Underwood, B. J., & Schulz, R. W. *Meaningfulness and verbal learning.* Philadelphia: Lippincott, 1960.

Webb, L. W. Transfer of training and retroaction: A comparative study. *Psychological Monographs*, 1917, **24**(3, Whole No. 104).

Wichawut, C., & Martin, E. Selective stimulus encoding and overlearning in paired-associate learning. *Journal of Experimental Psychology*, 1970, **85**, 383–388.

Wichawut, C., & Martin, E. Independence of A–B and A–C associations in retroaction. *Journal of Verbal Learning and Verbal Behavior*, in press.

Wickens, D. D. Encoding categories of words: An empirical approach to meaning. *Psychological Review*, 1970, **77**, 1–15.

Williams, R. F., & Underwood, B. J. Encoding variability: Tests of the Martin hypothesis. *Journal of Experimental Psychology*, 1970, **86**, 317–324.

(Received October 20, 1970)

Psychological Review
1971, Vol. 78, No. 4, 333–337

THEORETICAL NOTES

SELF-PUNITIVE BEHAVIOR—MASOCHISM OR CONFUSION?[1]

PAUL DREYER AND K. EDWARD RENNER[2]

University of Illinois

The phenomenon of self-punitive behavior in rats was considered and experimentally demonstrated with human *S*s. It had been suggested in the previous work that by choosing to run to a painful shock, the rats were showing a preference for punishment, thus providing an analogue for masochism. This extension to masochism was considered inappropriate, however, because of the use of the concept of choice and the attribution of a state of preference to the *S* without an adequate preference methodology. The apparent paradoxical effects of punishment seen in self-punitive behavior were resolved by making a conceptual distinction between instrumental learning about response-contingent outcomes (expectancy) and motivational states (preference).

Recent studies by Brown, Martin, and Morrow (1964) and Melvin, Athey, and Heasley (1965) have demonstrated that under certain conditions exposure to shock in a straight alley during extinction will retard the progress of extinction. Specifically, rats trained to run to escape shock which is present in the start box and alley will continue to run when the shock is present only in the alley, unlike other animals which are given regular unpunished extinction trials. This phenomenon has been labeled self-punitive behavior.

CURRENT EXPLANATIONS OF SELF-PUNITIVE BEHAVIOR

There are two explanations for self-punitive behavior current in the literature. The first (Church, 1963) is based on the discrimination hypothesis which states that the rate of extinction depends on the extent to which the conditions of extinction are discriminably different from the conditions of acquisition. The second is the vicious circle hypothesis originally proposed by Mowrer (1947) in which the punishment maintains the fear which motivates the response which is punished.

The Discrimination Hypothesis

In the normal punishment-facilitation experiment, the start box is above the runway. At the start of a trial, a warning signal comes on followed by the opening of a trap door which drops the rat onto the electrified grid floor of the runway below. According to the discrimination hypothesis, rats punished during extinction spend so little time in the runway area under the start box that there is no opportunity for them to learn that the first part of the runway no longer delivers shock. In contrast, subjects in the regular-extinction group are not shocked in the runway; thus, these subjects run more slowly, allowing discrimination and the subsequent extinction to occur.

The explanation based on the discrimination hypothesis has been criticized (Brown, 1969) in that discrimination is defined by whether or not the rat runs without providing independent evidence as to whether or not the change in conditions has been discriminated. In order to avoid circularity, it would be necessary to show that the subjects punished during extinction had not discriminated the change, but that rapid extinction occurred when the change was discriminated.

Another criticism of the discrimination hypothesis (Brown, 1969) is that it fails to account for the results of studies by Melvin (1964) and Martin and Moon (1967). These studies showed that groups of rats trained under 33% and 67% shock schedules and then shifted to 100% shock during extinction were just as resistant to extinction as groups maintained on their original schedules. Brown interpreted the discrimination hypothesis to require that a schedule shift would facilitate extinction. However, the discrimination hypothesis does not require that any change whatsoever in the situation will facilitate ex-

---

[1] Supported by Grant MH 11633 from the National Institute of Mental Health, United States Public Health Service.

[2] Requests for reprints should be sent to K. Edward Renner, Psychology Building, University of Illinois, Champaign, Illinois 61820.

tinction. The rate of extinction should be influenced only by changes which do in fact affect a discrimination between shock and nonshock in the startbox.

The Vicious Circle Hypothesis

The vicious circle hypothesis, which is favored by Brown (1969), has four basic assumptions: (*a*) Fear becomes conditioned to environmental cues during training because of contiguity. (*b*) Fear of the start box provides the motivation for running. (*c*) Running is reinforced by fear reduction in the goal box. (*d*) The acquired fear-arousing properties of the situation are strengthened on each self-punitive trial in which shock is encountered in the runway. Brown (1969) concludes that "the rats run because they are afraid and continue to be afraid because they encounter punishment when they run [p. 503]."

There is no need however, to assume that the two hypotheses are actually in conflict. The discrimination hypothesis states that to the degree to which rats fail to discriminate a safe from an unsafe start box they will continue to run. The vicious circle hypothesis explains from the point of view of the experimenter why this is so. It does not matter for the vicious circle hypothesis whether or not shock is present as long as fear is elicited in the start box. When fear is present, the subjects punished during extinction will then be subjected to further shock and caught in the circle, while regular-extinction subjects will extinguish to the fear-producing cues in the start box.

The distinctive point about a vicious circle is that the behavior should show unusual persistence over a long period of time, assuming other conditions do not change. Given the correctness of the vicious circle hypothesis, the eventual extinction which does occur indicates that subjects punished during extinction do learn that there is no shock in the area under the start box. But if this is true, it implies that when extinction started, the rats punished during extinction were not discriminating the change in the area under the start box, or at least could not inhibit taking the first few steps down the runway. The unanswered problem is, therefore, why discrimination or suppression of running is *slower* for the punished-extinction as opposed to the regular-extinction rats.

EXTENSION OF SELF-PUNITIVE BEHAVIOR TO THE CONCEPT OF MASOCHISM

Aside from arguing against the discrimination hypothesis and for the vicious circle formulation, Brown (1969) assumes that the rats which are punished, in comparison to the rats receiving nonpunished extinction, are *choosing* to leave the safe area under the start box to run across a shock grid to a safe goal box. He writes: "although the start box is conventionally described as the safe section of the runway, the fact that the self-punitive rat repeatedly chooses to expose himself to shock rather than remain in the start box permits one to define the start box as more aversive than the shock or the shock as more attractive than the start box [p. 506]." It is this extension of the data, not the data themselves, which requires Brown to reject the discrimination hypothesis. The attribution of choice to his animals requires, by definition, that they have discriminated between the alternatives available to them. Only by excluding the discrimination hypothesis can the extension of self-punitive behavior to masochism be maintained as an appropriate analogue.

In Brown's (1965) conception of masochism, he argues that the attractiveness or the aversiveness of physical stimuli can be determined only by the behavior of the organism. Events which are approached are called attractive, while those which are avoided are called aversive. According to this "behavioral criterion," masochistic behavior occurs when an individual approaches an event usually avoided by other individuals or by the same individual at different times. Thus rats punished during extinction are masochistic in that they approach electric shock which other rats (and they themselves later) normally avoid.

At first appearance, Brown's "behavioral criterion" may seem to be a useful way of looking at the problem of masochism, but if it is applied, then all at once many behaviors which in common usage are not considered masochistic will become so. Is the man who steps off a curb into an oncoming car (something he usually avoids) being masochistic? Clearly, observing his behavior is not enough to give this information (unless he is declared masochistic by fiat). It is necessary to know something about the state of his expectancies and preferences when he steps off the curb. Does the evidence establish, as Brown suggests, that rats receiving punishment during extinction are masochistic, that is, that they "choose" to run over shock to the goal box rather than remain in the start box?

In order to answer this question, the concept of choice must be considered. First it must be established that the start box is perceived by the rats as not containing shock. As an illustration of this requirement, consider the following hypothetical illustration. If rats were trained in a T maze containing 5 grams of food in the left arm and nothing in the right arm, one would expect consistent selection of the rewarded arm. If 10 grams of food were then placed in the right arm, we would not be justified in claiming that 5 grams were preferred over 10 grams because the rat continued to go to the left arm. The statement of preference could only be made if the animals "knew" that 10 grams were in the right arm. Not only must at least two discrete outcomes exist, but they must be discriminated by the organism making the choice. The preference which choice implies does not exist when one of the behaviors has a consequence which is unknown to the subject.

In most punishment-facilitation studies, the starting point is one of the alternatives and there is no way of knowing, a priori, that not running exists as a nonshock alternative for the rat. When requirements for a concept of choice are considered, Brown's extension to masochism is no less circular than the discrimination hypothesis because there is no independent evidence that a choice exists.

A nonpunished control group is included in Brown's research in order to have a standard against which to compare the behavior of the rats punished during extinction. Brown's procedure, however, establishes only that two groups discriminate these alternatives at a different rate, the understanding of which is an interesting psychological question. The empirical fact of this differential rate is meaningless, however, for an extension to masochism, because the methodological procedures of the usual punishment-facilitation studies do not permit the use of the concept of choice and preference. The conceptual difficulty can be illustrated by a parallel experiment conducted with humans.

An Illustrative Human Experiment

Method. The subjects were seated alone in a room before a numerical readout screen. At the start of a trial, the experimenter said "Get ready" over an intercom, one to two seconds later the number 01 appeared on the screen and a buzzer sounded for .5 seconds, followed immediately by a pulsating shock to the index and ring finger of the nondominant hand. The dominant hand was free to operate a telegraph key. The subjects were instructed that:

> The electric shock may be escaped by pressing the telegraph key. Twenty presses on the key are required to increase the number on the screen by one. Once you get the screen to count up to the number ten, the shock will turn off.

After five acquisition trials, during which the shock was successively increased to a maximum of 2.5 milliamperes, the subjects, all of whom were male, were alternately assigned to one of two extinction groups. For the regular-extinction group, the shock was turned off. For the punished-extinction group, the first key press after the buzzer turned on the shock, which could be escaped only by advancing the screen to the number 10 (180 presses). Both groups received five extinction trials. The time required to traverse the "runway" by key pressing was recorded. A time limit of 50 seconds, which was set on all trials, also served as the extinction criterion.

It was expected that subjects would start to make key press responses during the warning period or at least pounce on the key as soon as the buzzer stopped. Once the extinction phase of the experiment started, the regular-extinction group should soon discover that the shock had been terminated and stop responding or slow down. The second group, which paralleled Brown's punished-extinction group, should find only that they continue to be punished and must press the key to escape the punishment. All subjects punished during extinction who had not discovered at the end of the five extinction trials that it was their own key pressing which was causing them to be shocked were instructed as to the actual contingencies and given an additional trial. All subjects filled out a standard questionnaire and were interviewed at the end of the experiment.

Results. The results are shown in Figure 1. Except for the lack of a learning curve during acquisition, the figure is indistinguishable from the animal data on self-punitive behavior. The acquisition phase of key pressing is close to the capacity of the subjects, reflecting that humans know how to press a key rapidly. The apparent separation of the two groups during acquisition was not significant ($F = 1.86$, $df = 1/20$) and reflected only that the regular-extinction group contained several subjects who could not generate fast key pressing; there was no qualitative difference in their behavior. The extinction data are of interest on two counts: First, the group punished during extinction pressed the key at a significantly faster rate than the regular-extinction group ($F = 11.49$, $df = 1/20$, $p < .01$) over the five extinction trials. Second, 10 of

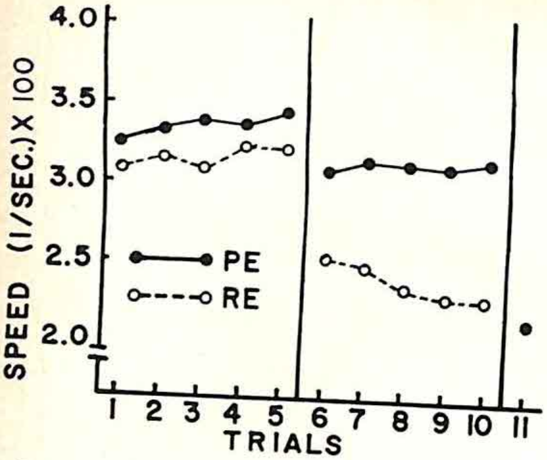


FIG. 1. Speed with which 180 telegraph key presses were completed on the five acquisition and five extinction trials for the punished-extinction (PE) and regular-extinction (RE) subjects. (Subjects punished during extinction received an eleventh trial after they were informed that it was their initial key press which turned on the shock. The 50-second limit on each trial imposed a lower boundary of 2.0 on the speed score. The maximum speed obtained by any individual subject was 4.0.)

the 11 subjects in the group punished during extinction stopped pressing the key once they were informed of the contingencies (Trial 11 of Figure 1).

The questionnaire and interview data served to confirm what is intuitively obvious. Nine of the 11 subjects punished during extinction continued to press for all five extinction trials and reported that they did not know the shock was being turned on by their own key presses. Two of the 11 subjects punished during extinction always waited for the shock to begin during acquisition and immediately detected the change at the start of extinction, refraining from making any responses during extinction. Six regular-extinction subjects pressed the key for all the five extinction trials, but all did so at a slower rate than during acquisition. Of the six, two reported that they thought the shock might come on again so they wanted to be close to the end, while the other four reported that they wished to get the experiment over with. The five regular-extinction subjects who stopped pressing altogether stated that there was no need to press because there was no shock. Nine of the 11 regular-extinction subjects made key-press responses at the start of the sixth trial and presumably would not have detected the change had they been in the punished-extinction condition.

The human data are quite similar to the animal data, in which some subjects punished during extinction detect the change and immediately stop responding, and some regular-extinction subjects continue to run for a considerable period of time, but at a reduced rate (see Brown, 1969).

Discussion. What is obvious in the human situation also applies conceptually to the animal situation. The experimental procedure simply does not permit the introduction of the terms "choice," "preference," or "masochism." What the procedure shows is that subjects who find themselves on a charged grid, run (press) faster than those subjects who are not on a charged grid. As the nonpunished subjects learn that the shock has been removed, they slow down, and thus the alternative response of not running (not pressing) is acquired. For those subjects receiving punishment, the no-shock outcome is outside the range of response variation and therefore does not exist as an alternative; thus the concepts of choice and preference do not apply.

Requirements for an Extension to Masochism

Our human experiment and the discussion of the animal work on self-punitive behavior are intended to illustrate that there is no need at the present time to cast the discrimination hypotheses and the vicious circle hypothesis into the role of competing alternative explanations. Clearly, fear motivates responding, and a change in behavior does not occur until a discrimination between the alternatives has taken place. Our evidence with humans suggests that as soon as this discrimination takes place, subjects punished during extinction will stop responding. In experiments on self-punitive behavior, it is necessary to separate *expectancy* questions about response-outcome contingencies from motivational questions about *preferences* between those outcomes. The two questions require different methods and both, it would seem, are relevant for an adequate consideration of self-defeating and self-punitive behavior.

There is no basis at this time to infer choice or preference for running (or pressing) over not running (or not pressing) on the part of subjects punished during extinction, or to suppose that the phenomenon of self-punitive behavior provides an analogue for masochism, as long as the term is used to imply the preference for a painful event. The "behavioral criterion" proposed by Brown (1965, 1969) is inadequate for it does not use a preference methodology (e.g., Irwin, 1961).

Our critique of the existing research is *not* intended to dispute the reliability of the empirical phenomenon of self-punitive behavior. Indeed, it is as reproducible with humans as with animals. But, the roles of fear and discrimination both need to be considered. Whether the basic paradigm can be used to provide an analogue for masochistic behavior yet remains to be demonstrated through the use of preference methodology. To meet the constraints of a preference methodology, it would be necessary to show that running into the shock by the punished-extinction subjects is preferred over some alternative; but that the regular-extinction subjects prefer (or are indifferent to) the same alternative versus running without shock. Such a demonstration seems unlikely.

REFERENCES

Brown, J. S. A behavioral analysis of masochism. *Journal of Experimental Research in Personality*, 1965, **1**, 65–70.

Brown, J. S. Factors affecting self-punitive locomotor behavior. In B. A. Campbell & R. M. Church (Eds.), *Punishment and aversive behavior*. New York: Appleton-Century-Crofts, 1969.

Brown, J. S., Martin, R. C., & Morrow, M. W. Self-punitive behavior in the rat: Facilitative effects of punishment on resistance to extinction. *Journal of Comparative and Physiological Psychology*, 1964, **57**, 127–133.

Church, R. M. The varied effects of punishment on behavior. *Psychological Review*, 1963, **70**, 369–402.

Irwin, F. W. On desire, aversion, and the affective zero. *Psychological Review*, 1961, **68**, 293–300.

Martin, R. C., & Moon, T. Self-punitive behavior and periodic punishment. Paper presented at the meeting of Psychonomic Society, Chicago, October 1967.

Melvin, K. B. Escape learning and "vicious circle" behavior as a function of percentage of reinforcement. *Journal of Comparative and Physiological Psychology*, 1964, **58**, 248–251.

Melvin, K. B., Athey, G. I., Jr., & Heasley, S. H. Effects of duration and delay of shock on self-punitive behavior in the rat. *Psychological Reports*, 1965, **17**, 107–112.

Mowrer, O. H. On the dual nature of learning—a re-interpretation of "conditioning" and "problem-solving." *Harvard Educational Review*, 1947, **17**, 102–148.

(Received September 4, 1970)

1971, Vol. 78, No. 4, 338–340

STRENGTH THEORY AND CONFIDENCE RATINGS IN RECALL

HARLEY A. BERNBACH[1]

Cornell University

A Monte Carlo simulation of strength theory and confidence ratings was run, because a previous argument, that strength theory cannot account for the observed invariance of Type 2 receiver operating characteristic curves with respect to the probability of a correct response, was incorrect. It is concluded, as before, that strength theories relating confidence to strength cannot predict this invariance.

Two different types of theories about the nature of the memory trace have been applied to confidence ratings in short-term memory. One, a strength theory (e.g., Wickelgren & Norman, 1966), effectively postulates a continuous distribution of memory states. The other postulates only two states. Both can handle recognition memory data about equally well. But Murdock (1966) reported an unexpected invariance in one measure of a subject's performance in recall experiments: the Type 2 receiver operating characteristic (ROC) curve does not vary with serial position. This is an important finding, for it may allow us to distinguish between the two classes of theories. The two-state theory clearly predicts this result (Bernbach, 1967). But until now, it has not been possible to derive the predictions for the continuous strength theory.

The purpose of this note is to present the predictions from continuous strength theory and to show that they are incompatible with the data presented by Murdock (and replicated by Bernbach, 1967, 1970; Bernbach & Bower, 1970). This note therefore modifies my 1967 statement about continuous strength theory. The reasoning in that paper was incorrect. This note, however, shows that when the predictions of continuous strength theory are worked out (by means of a Monte Carlo simulation), the prediction of the earlier paper is correct, even though it was made then for improper reasons.

The Simulation

First it was necessary to make some assumptions about the relationship between confidence judgments and a strength theory for a multialternative recall task. The strength theory adopted has normally distributed strength distributions and uses the traditional "habit family hierarchy" theory, which assumes that the response emitted is the one with the highest strength at the instant of recall. In the absence of specific theory relating confidence and strength, two basic types of model for confidence were tried: an absolute strength model and a relative strength model. The first model assumed confidence to be a function of the value of the strength of only the emitted response, while the second assumed confidence to be a function of the difference in strength between the two highest items at the instant of recall. It was felt that these models represented end points on a continuum of possible models for the relationship between strength and confidence.

For the Monte Carlo runs, a task with just four response alternatives was assumed, mainly to keep down the computer time used. Empirically, such a task was discussed by Bernbach (1967, p. 477). At that time, the ROC curves shown were modified to "correct for guessing," so that the two-state model would predict no change across trial number. This was necessary because the two-state model predicts invariance in ROC curves with respect to the states, and the hit rates were based on the correct responses, which included some correct guesses from the guessing state.

It was these modified ROC curves that were shown to be invariant with respect to trial number. Thus, the same modification for guessing was applied to the simulated data before ROC curves were plotted to investigate whether differences were predicted by the continuous strength models. It should be noted that this modification has the property of bringing the curves closer together and, if anything, works against the thesis of this note.

The simulation proceeded as follows. A complete run was made up of 5,000 random draws. On each, four numbers, representing strength values, were drawn at random from normal

---

[1] Requests for reprints should be sent to Harley A. Bernbach, Department of Psychology, Morrill Hall, Cornell University, Ithaca, New York 14850.

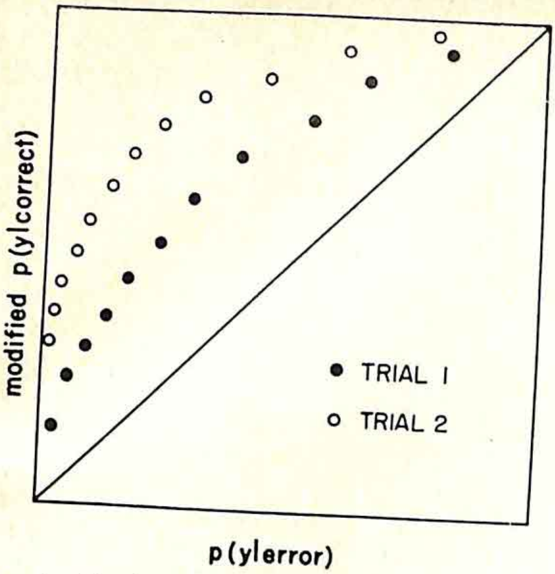


FIG. 1. Simulated and modified-for-guessing ROC curves for the relative confidence model.

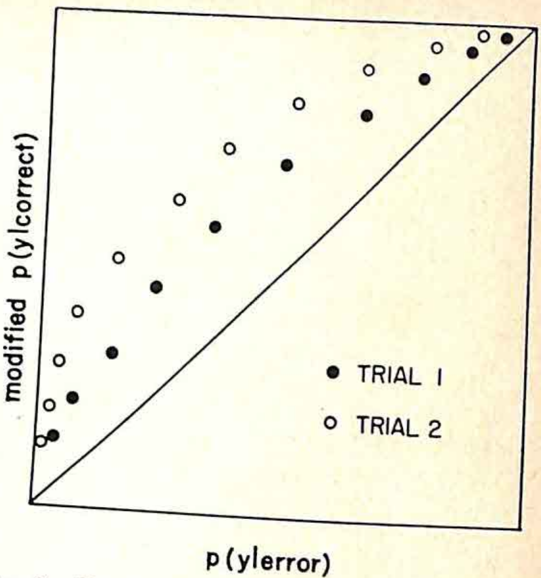


FIG. 2. Simulated and modified-for-guessing ROC curves for the absolute confidence model.

distributions, all with standard deviations equal to one. This is the simplest case of equivariate strength distributions and is most closely related to the strength models for recognition memory proposed by Wickelgren and Norman (1966). The first distribution was designated "correct," and depending on trial number (designating the two ROC curves to be plotted), had a mean of 1 or 2. These values were chosen arbitrarily to give some range to the probability correct (which turned out to be .55 and .83, respectively). The remaining three distributions were designated as errors, and their means were zero.

In the computer program, an array with two rows and 15 columns was established for each model, relative confidence and absolute confidence. If the highest of the four numbers generated at random from the four distributions was the first, a number in one of the cells in the first row of each array was incremented by one; otherwise, a number in a second row cell was incremented. The columns of the cells incremented depended on the confidence model involved. An example will show how this was done.

Suppose the four values generated in a draw were 1.37, .26, −.43, and .03. Since the first is the highest, a number in the first (or "correct") row of each array would be incremented by one. For the absolute model, the number 1.37 would be multiplied by 4 (a multiplier chosen arbitrarily to provide some range of ROC points) and rounded off to 5. Thus, the number in cell 1 by 5 would be incremented by one. For the relative model, the difference between the two highest numbers, 1.11, would be multiplied by 6 (again an arbitrary multiplier) and rounded off to 7. In the array for the relative model, cell 1 by 7 would be incremented. The ROC curves were obtained from the filled arrays in the standard manner, and then were modified for guessing, using a guessing rate of one-fourth.

The simulated (and modified-for-guessing) ROC curves for the relative strength model are shown in Figure 1. It is quite clear that the distance between the modified ROC curves and the major diagonal increases with the probability of a correct response. This is also true of the absolute strength model, as shown in Figure 2. As mentioned earlier, the empirical evidence shows no such change, even with larger differences in correct response probability than that obtained from this particular Monte Carlo simulation. (Several alternative simulations have also been carried out, each with the same result: the continuous strength theory cannot predict invariant Type 2 ROC curves. Among the variations tested were different rules for generating confidences and allowing the means and variances of the strength distributions to differ.)

Discussion

Of course, it is not possible to generalize the above results to every theory of the memory trace that relates a strength continuum to confidence. The models considered were reasonable ones, however, and the results quite clear. Despite the faulty earlier argument (Bernbach, 1967), it would seem, until a model to the contrary appears, that the invariance of Type 2 ROC curves relative to recall probability argues for the two-state model.

It should be made clear that the preceding argument is addressed only to strength theories that, like the models of Wickelgren and Norman (1966), postulate a direct relation between strength and confidence. A strength theory like that proposed by Shiffrin (1970) is, in fact, a two-state theory for confidence ratings. In Shiffrin's search theory, an item is in either a *found* state or a *not found* state at the time of test. Though the strength of an item determines which state it is in, confidence can be based on distributions for only these two states. Thus, the *underlying* theory of the memory trace need not be finite state. The theory must provide, however, that as far as the subject's awareness and ability to rate confidence are concerned, there are two and only two states. As stated in the earlier paper (Bernbach, 1967), there are a multitude of specific theories of memory that are compatible with this.

REFERENCES

BERNBACH, H. A. Decision processes in memory. *Psychological Review*, 1967, **74**, 462–480.

BERNBACH, H. A. Recognition and recall in short-term memory. Paper presented at the meeting of the Psychonomic Society, San Antonio, November 1970.

BERNBACH, H. A., & BOWER, G. H. Confidence ratings in continuous paired-associate learning. *Psychonomic Science*, 1970, **21**, 252–253.

MURDOCK, B. B., JR. The criterion problem in short-term memory. *Journal of Experimental Psychology*, 1966, **72**, 317–324.

SHIFFRIN, R. M. Memory search. In D. A. Norman (Ed.), *Models of human memory*. New York: Academic Press, 1970.

WICKELGREN, W. A., & NORMAN, D. A. Strength models and serial position in short-term recognition memory. *Journal of Mathematical Psychology*, 1966, **3**, 316–347.

(Received October 15, 1970)

Psychological Review
1971, Vol. 78, No. 4, 341–343

THE BEHAVIORISTIC CONCEPTION OF NEUROSIS:
A REPLY TO TWO CRITICS

JOSEPH WOLPE[1]

Temple University Medical School

Criticisms of the behavioristic view of neurosis by Costello and by Weitzman are answered. It is observed that some of the criticisms arise from failure to note the centrality of autonomic responses.

The idea that neurotic behavior consists of conditioned responses of specifiable character has in the past decade been the subject of a considerable amount of experimental and clinical research that has largely supported it (see, e.g., Paul 1969). This paper replies to two recent critiques by Weitzman (1967) and by Costello (1970). Since Weitzman's criticisms are mainly in the clinical area, while Costello's deal essentially with learning mechanisms, it will be logical to discuss the latter first.

Costello

Theory of neurosis. (*a*) Much of Costello's criticism stems from his erroneous belief that behaviorists regard neurosis as an instance of *motor avoidance conditioning.* In fact, the definition that has been generally adopted is this: A neurosis is a *persistent unadaptive habit that has been acquired by learning in one or a series of anxiety-generating situations. Conditioned anxiety is usually the central component of neurotic habits* (Wolpe, 1958). Anxiety is defined as an individual organism's characteristic constellation of autonomic responses to noxious stimulation. It is ironic that in referring to Eysenck's recently expressed adoption of a classical aversive conditioning theory rather than an avoidance one, Costello states, "it does not appear to have resulted in any proposed changes in the understanding and methods of treatment of phobias and need not be discussed further [p. 251]." The methods of behavior therapy have always been based on the classical conditioning theory implied in the above definition of neurosis.

(*b*) Costello's contention that animals do not usually learn to avoid the conditioned stimulus to an avoidance reaction is obviously irrelevant to this definition of neurosis. What is relevant is that that stimulus, in the course of the same set of conditioning operations that conditioned it to the avoidance reaction, will also have become conditioned to the *anxiety* responses that the unconditioned stimulus evoked. Evocations of the conditioned anxiety response are an effective source of second-order conditioning of that response (Wolpe, 1952a; 1958, p. 60)—in contrast to evocations of an avoidance response (McAllister & McAllister, 1964).

(*c*) Costello declares that "the explanation usually accepted by behavior therapists" for the resistance to extinction of phobias is Solomon and Wynne's (1954) proposition that extinction is prevented because the avoidance response makes "reality testing" (by the unreinforced presentation of the CS) impossible. This explanation is at odds with the facts of experimental neurosis (Wolpe, 1952a; 1958, p. 71). Subjecting experimentally neurotic animals to repeated or prolonged exposure to the stimuli eliciting the neurotic responses does not lead to extinction. In other words, when "reality testing" does occur, it is not "curative." What might account for the resistance to extinction is that removing the animal from the experimental cage and its anxiety-evoking stimuli results in anxiety-drive reduction, which reinforces the anxiety responses to the cage stimuli that were evoking them (Wolpe, 1952a, 1952b).

Causation. (*a*) Costello states that although Wolpe (1952a) and Smart (1965) demonstrated that no conflict is necessary to produce experimental neuroses, "even the most casual observation of clinical phobias" suggests that conflict is involved in them. He adduces no evidence to support this statement. The ubiquity of conflict in clinical neuroses is not borne out by patients' histories. Most phobias, as well as "more complex" anxiety habits, have no discernible relation to any ongoing conflict; nor need conflict have played any part in their origin. For example, a driving phobia may date from a panic occasioned by an attack of hypoglycemia during driving. On the other hand, of course, conflict *can* be the "unconditioned" source of the anxiety response to which neurotic conditioning occurs (Wolpe, 1958, pp. 38 and 79).

---

[1] Requests for reprints should be sent to Joseph Wolpe, Department of Psychiatry, Temple University Medical School, c/o Eastern Pennsylvania Psychiatric Institute, Henry Avenue, Philadelphia, Pennsylvania 19129.

(*b*) The fact that Bregman (1934) failed to reproduce the phobia that Watson and Reyna had induced in Little Albert has little significance. In the first place, people are not all equally conditionable. Second, failure was probably ensured by the fact that "more neutral stimuli" were used as potential conditioned stimuli. Seligman (1970) has shown how relevant the character of a stimulus is to its conditionability to a particular response.

(*c*) Costello argues that the success of systematic desensitization does not support the learning theory of phobias because "methods other than systematic desensitization appear to be equally successful." The equal success of other methods has by no means been demonstrated (see, e.g., Nawas & Mealiea, 1971; Paul, 1966; Sushinsky & Bootzin, 1970; Willis & Edwards, 1969). To the extent that methods derived from a causal theory are egregiously successful, the theory becomes more plausible; but direct testing in the causal context can alone *establish* a causal theory.

(*d*) Costello finds the suggestion of Andrews (1966) that phobias develop in relation to the satisfaction of dependency needs "at least equally plausible, despite the lack of an experimental analogue." We are not told what considerations make it plausible. The conditioning theory of the development of phobias at present looks quite persuasive—it has an experimental base, it is borne out by the details patients give of the onset of their fears, the clinical methods that are its fruit have been attended by success, and no factual considerations have been brought forward that seriously challenge it.

Weitzman

Weitzman states that the major substance of his essay is "an examination of the grounds on which clinicians have based their rejection of behavior therapy."

The relation of "learning theory" to behavior therapy. On the basis of a consensus among experimental psychologists (Estes, Koch, MacCorquodale, Meehl, Mueller, Schoenfeld, & Verplanck, 1954) that "there is no monolithic system properly called learning theory," Weitzman deduces that there is no learning theory competent to handle the data of either behavior therapy or psychotherapy. Behavior therapists may give a misleading impression when they speak of behavior therapy arising out of "learning theory," but their allusion is not to any global theoretical system, but to particular principles and paradigms that experiments in the psychological laboratory have yielded (see, e.g., Skinner, 1953; Wolpe, 1958). An experimentally established paradigm is a solid possession which behavior therapists apply to clinical problems. This enterprise has borne results that support the hypothesized parallelism between the laboratory and the clinic.

Weitzman's criticisms of systematic desensitization. (*a*) Weitzman states that the derivation of systematic desensitization from the treatment of animal neuroses by systematic feeding in the presence of increasingly strong anxiety-evoking stimuli hides "a form of analogy-making which gives comfort to the behavior therapist but which obscures differences of profound significance between the systematic desensitization of cats and men [p. 304]." Surely, if parallel procedures in cats and men cure the neuroses of both, the species differences must be judged irrelevant in the particular context.

(*b*) In the standard desensitization procedure, the scene the patient is directed to imagine is regarded as paralleling the exogenous stimulus situation in which the experimental cat is desensitized. Weitzman endorses the view of Breger and McGaugh that the use of the term "stimulus" is here only "remotely allegorical" to the traditional use of the term in psychology. There are two good reasons for rejecting this contention. First, on Lundberg's (1937) definition of "stimulus" as "that to which response is made," a succession of words or a visual image is just as much a stimulus in relation to the response that follows it as an exogenous stimulus would be. Second, several experiments have shown that much the same desensitizing results are found if extrinsic stimuli are used in the place of imaginal stimuli (e.g., Cooke, 1966; Goldberg & D'Zurilla, 1968).

(*c*) Weitzman describes an experiment he undertook on six patients to investigate their experiences during the silent period when they were supposed to be imagining hierarchical scenes. He states that they all reported a shifting flow of visual imagery which he feels could provide "a free association in fantasy." However, he had them maintain "static visualization of the scene described" for periods of up to one minute—far more than the standard duration of 5–7 seconds for scene presentations in systematic desensitization (Wolpe, 1969). Patients seem almost invariably able to maintain constant visualizations for the standard short spells.

Weitzman's defense of psychoanalytic psychotherapy. Weitzman tries to rescue psychoanalytic theory from the implications of the failure of symptom substitution to occur after behavior therapy by putting forward three separate proposals. The first of these is that symptoms may disappear permanently through means other than the breakdown of repression, quoting Freud's (1936) distinction "between mere repression and the true disappearance of an old desire or impulse [p. 83]." Weitzman does not suggest by what mechanism consonant with psychoanalytic theory desensitization could achieve the lasting disappearance of "old desires and impulses."

The second proposal is that somehow desensitization may be a repression-lifting mechanism. There would need to be evidence that presumed repressed material does somehow emerge in the course of desensitization—and of this, those who routinely use the technique have so far seen no sign.

Weitzman's third proposal is that systematic desensitization, by diminishing fear, increases ego strength, and may thus signal the possibility of creating more adequate "binding behaviors." "Why then," he asks, "predict symptom substitution?" But can the patient be safe from it as long as the repressed energies remain in thrall? And does not this proposal undercut some other tenets of psychoanalytic theory outlined by Weitzman? If anxiety "serves as a signal to the ego that a dangerous instinctual demand is growing in strength" and if an external phobic stimulus must "symbolically represent the meaning of the internal danger," it follows that the removal of anxiety by desensitization must deprive the patient of his valuable warning system without removing the "internal danger."

REFERENCES

ANDREWS, J. D. W. Psychotherapy of phobias. *Psychological Bulletin*, 1966, **66**, 455–480.

BREGMAN, E. O. An attempt to modify the emotional attitudes of infants by the conditioned response technique. *Journal of Genetic Psychology*, 1934, **45**, 169–198.

COOKE, G. The efficacy of two desensitization procedures: An analogue study. *Behaviour Research and Therapy*, 1966, **4**, 17–24.

COSTELLO, C. G. Dissimilarities between conditioned avoidance responses and phobias. *Psychological Review*, 1970, **77**, 250–254.

ESTES, W. K., KOCH, S., MACCORQUODALE, K., MEEHL, P. E., MUELLER, C. G., JR., SCHOENFELD, W. N., & VERPLANCK, W. S. *Modern learning theory*. New York: Appleton-Century-Crofts, 1954.

FREUD, S. *The problem of anxiety*. New York: Norton, 1936.

GOLDBERG, J., & D'ZURILLA, T. G. Demonstration of slide projection as an alternative to imaginal stimulus presentation in systematic desensitization therapy. *Psychological Reports*, 1968, **23**, 527–533.

LUNDBERG, G. A. *Foundations of sociology*. New York: Macmillan, 1937.

MCALLISTER, D. E., & MCALLISTER, W. R. Second order conditioning of fear. *Psychonomic Science*, 1964, **1**, 383–384.

NAWAS, M. M., & MEALIEA, W. L. Systematic desensitization vs. implosive therapy. *Journal of Behavior Therapy and Experimental Psychiatry*, 1971, in press.

PAUL, G. L. *Insight versus desensitization in psychotherapy*. Stanford: Stanford University Press, 1966.

PAUL, G. L. Outcome of systematic desensitization I. and II. In C. M. Franks (Ed.), *Behavior therapy: Appraisal and status*. New York: McGraw-Hill, 1969.

SELIGMAN, M. E. P. On the generality of the laws of learning. *Psychological Review*, 1970, **77**, 406–418.

SKINNER, B. F. *Science and human behavior*. New York: Macmillan, 1953.

SMART, R. G. Conflict and conditioned aversive stimuli in the development of experimental neuroses. *Canadian Journal of Psychology*, 1965, **19**, 208–223.

SOLOMON, R. L., & WYNNE, L. Traumatic avoidance learning: The principles of anxiety conservation and partial irreversibility. *Psychological Review*, 1954, **61**, 353–385.

SUSHINSKY, L. W., & BOOTZIN, R. R. Cognitive desensitization as a model of systematic desensitization. *Behaviour Research and Therapy*, 1970, **8**, 29–33.

WEITZMAN, B. Behavior therapy and psychotherapy. *Psychological Review*, 1967, **74**, 300–317.

WILLIS, R. W., & EDWARDS, J. A. A study of the comparative effectiveness of systematic desensitization and implosive therapy. *Behaviour Research and Therapy*, 1969, **7**, 387–396.

WOLPE, J. Experimental neurosis as learned behavior. *British Journal of Psychology*, 1952, **43**, 243–268. (a)

WOLPE, J. Formation of negative habits: A neurophysiological view. *Psychological Review*, 1952, **59**, 290–299. (b)

WOLPE, J. *Psychotherapy by reciprocal inhibition*. Stanford: Stanford University Press, 1958.

WOLPE, J. *The practice of behavior therapy*. New York: Pergamon Press, 1969.

(Received December 4, 1970)

Psychological Review
1971, Vol. 78, No. 4, 344–347

AVOIDANCE CONDITIONING AND BEHAVIOR THERAPIES: A REPLY TO COSTELLO

ARNOLD POWELL[1] AND AUGUSTUS R. LUMIA[2]

University of Nebraska

Costello's conclusion that avoidance responses and phobias are dissimilar was criticized on several points: (*a*) Costello had proposed that avoidance responses are adaptive and phobias unadaptive. However, behaviors cannot be functionally distinguished on the basis of adaptiveness, since adaptiveness refers to a relationship between a given response and particular environmental events. (*b*) Given Costello's definition of adaptiveness, it can be shown that avoidance behaviors are unadaptive under a variety of conditions. (*c*) Herrnstein's analysis of avoidance conditioning does not eliminate fear reduction as a sufficient condition for avoidance. Although theoretical interpretations of avoidance conditioning must take Herrnstein's analysis into account, this does not necessarily have any implications about the similarity between avoidance responses and phobias.

The view that two-factor theory of avoidance conditioning provides an appropriate model of the acquisition and maintenance of phobic reactions was recently criticized by Costello (1970). While Eysenck and Rachman (1965) were singled out as examples of behavior therapists who maintain such a view, the criticisms were considered appropriate for any position that postulates that phobias are acquired according to the principles of Mowrer's (1947) theory of avoidance conditioning. Furthermore, these specific criticisms represented an indictment against the behavior therapies in general. By providing "more specific detailing of discrepancies between the data of learning studies and the practices and formulations of the behavior therapists [Costello, 1970, p. 250]" Costello had hoped to establish a more defensible basis than had previous writers (e.g., Breger & McGaugh, 1965; Weitzman, 1967) for criticizing the increasingly popular behavior therapies.

Given the present state of knowledge in our science as well as in our technology, it is necessary to carry on a constant evaluation of the rationale and practices of psychotherapy irrespective of whether the model for the therapy process stems from the learning laboratory or from any other perspective in the field of psychology. Our laboratory models as well as the practice of psychotherapy benefit from such evaluations and comparisons. However, it is also important to recognize that discrepancies between the behavior of laboratory subjects and persons observed in the clinic can arise from a variety of sources and the occurrence of such discrepancies does not necessarily demand that we judge them to be governed by *different* underlying principles. Behaviors which are topologically different, for example, can be functionally similar. The present writers feel that the conclusion that conditioned avoidance reactions and phobias are "dissimilar" is somewhat premature and that many subtleties involved in arriving at such a conclusion need to be further explored.

Costello's (1970) principal thesis was that conditioned avoidance responses do not provide an appropriate model of phobic reactions because

> (*a*) the avoidance responses are adequate (coping) behaviors and (*b*) they do not involve a conflict with approach behaviors and such a conflict appears to be characteristic of clinical phobias [p. 252].

There are a number of difficulties associated with attempts to distinguish between these behaviors on the basis of their "adequateness" or "adaptiveness" (Costello, 1970, pp. 251 ff.). First, according to the definition provided for "adaptiveness,"

> Responses learned in the usual avoidance procedure are adaptive because they enable the animals simply to avoid a noxious stimulus. . . . But phobias are maladaptive because they prevent the occur-

---

[1] During the preparation of this paper the first author was supported by a National Science Foundation Predoctoral Traineeship.

Requests for reprints should be sent to Arnold Powell, who is now at the Department of Psychology, Columbus College, Columbus, Georgia 31907.

[2] Now at Skidmore College.

> rence of behaviors desired by the individual . . . and/or desired by society . . . [p. 251],

there is much evidence in the literature to support the conclusion that under many circumstances, avoidance responses could be considered "unadaptive." What is perhaps more important is the fact that regardless of the specific connotations or denotations given the term *adaptiveness,* it is not typically used to describe some inherent characteristic of behavior. In general, adaptiveness refers to some relationship between a particular behavior, or class of behaviors, and selected environmental events. The judgment that some behavior is "adaptive" or "unadaptive" must be predicated on the demonstration that this relationship does, or does not, hold. However, since adaptive pressures and environmental contingencies can change rapidly over time and can vary for organisms at any particular point in time, this judgment can only be made with regard to some specified behavior of a given organism at a specified point in time.

In order to comment further on the proposition that avoidance behaviors are "adaptive" and phobias are "unadaptive," we shall first examine a couple of alternative implications of the definition that Costello provides for these terms and then proceed to examine some relevant experimental findings. Because of the factors discussed in the preceding paragraph, any conclusions reached about the adaptiveness of avoidance behaviors must be qualified by "given a particular set of conditions and organisms."

One implication of the above definition would be to call a behavior *unadaptive* if once the response has been acquired, the organism will persist in making it even after the original environmental contingencies have been removed or even reversed. It should be possible to find many examples of phobias that could be described as unadaptive in this sense. In fact, this is often characteristic of a phobia; that is, the organism encounters some situation that is highly aversive but later finds that situation to provide a different set of contingencies, and the earlier acquired tendencies prevent him from responding appropriately to the new set of contingencies. On the other hand, it is possible to find examples of the same phenomenon in laboratory investigations of avoidance conditioning. For example, in several of the classic studies by Solomon and Wynne (1954) on traumatic avoidance conditioning, dogs given only a very few exposures to strong electric shock continued to make avoidance responses throughout hundreds of extinction trials. When Solomon and Wynne (1953; 1954, pp. 364 ff.) provided some "therapy" by reversing the contingencies and making the presentation of shock now contingent on making the response that previously avoided shock, the speed of this instrumental reaction was actually found to increase. And in another context, Sidman, Herrnstein, and Conrad (1957) have shown that free, or noncontingent, shocks administered in extinction to an animal that has already acquired an avoidance response will lead to an increase in the rate of that avoidance response. Finally, in a series of recent investigations, Brown and his associates (Brown, 1969) have presented considerable evidence for self-punitive behaviors in particular escape-avoidance situations.

There are still other implications of the way in which Costello employed the term "adaptive." To consider briefly another aspect, a behavior should be considered unadaptive if it disrupts the organism's ongoing behavior in some other situation, particularly if the objective environmental contingencies do not require that the disruptive behavior even occur in other situations. No doubt there are many examples of instances in which phobic reactions have this kind of disruptive effect on other behaviors. However, what is more important in terms of Costello's conclusions is the fact that certain elements in avoidance conditioning can also have this disruptive effect—as evidenced by the voluminous literature on conditioned emotional responses (cf. Campbell & Church, 1969).

It should be clear from the above considerations that according to Costello's use of the terms "adaptive" and "unadaptive," there are a variety of conditions under which avoidance behaviors may be described as unadaptive.

Recent studies on the phenomenon of "learned helplessness" (see Maier, Seligman, & Solomon, 1969) can be used to further illustrate the problems in describing a behavior as adaptive or unadaptive. It has been found that the presentation of noncontingent shocks prior to exposure to an avoidance contingency will have a very disruptive effect on avoidance acquisition. Should the behavior that is being acquired during the noncontingent stage be described as "unadaptive?" An unqualified "yes" or "no" cannot be given to this question, for if the subject should never be exposed to the avoidance contingency, then this behavior can be viewed as adaptive. Since the shocks are presented noncontingently, any effort expended in trying to alter or prevent their occurrence is wasted effort. However, if the avoidance conditioning is forthcoming, then we might describe learned helplessness as being unadaptive, because under these conditions the animal is responding to a set of contingencies that are no longer present. The general point is that the "adaptiveness" of avoidance behavior as well as any other behavior including phobias can only be defined in terms of the particular response under a particular set of conditions.

It is interesting to note that the situation with Solomon and Wynne's (1953) subjects contains the element of conflict that Costello indicates is so characteristic of phobias observed in the clinic. That is, the dogs given traumatic avoidance conditioning were later presented with a conflict between "approaching" in order to avoid shock and actually jumping into shock. When presented with such a conflict, the subjects jumped even faster than before the element of conflict was introduced. This suggests the possibility that it is the presence of conflict that makes the behavior of an organism that has acquired a phobia seem so unadaptive. That is, the acquisition of strong fear or avoidance tendencies is not, in itself, unadaptive, but with reversal of contingencies that introduces conflict, even greater fear and tendencies to avoid are evoked. In spite of the fact that many organisms with phobic reactions also happen to have conflicts present, it is possible that avoidance behaviors and phobias are acquired according to the same principles and that Mowrer's two-factor theory specifies those principles.

Costello relied heavily on Herrnstein's (1969) recent interpretation of avoidance conditioning, according to which the CS in avoidance conditioning serves a discriminative function—its most salient characteristic being that it sets the occasion for the avoidance response. This is in contrast to Mowrer's two-factor theory in which CS onset arouses fear or anxiety that motivates the avoidance response, and CS offset provides the reinforcement (via fear reduction) for avoiding. Herrnstein was extremely ingenious in finding experimental avoidance tasks in which CS offset was not a necessary condition for avoidance to occur. For example, avoidance responses were acquired in situations where the only effect of that response was to reduce (for a given time interval) the future probability (frequency) of shock presentation (Herrnstein & Hineline, 1966). Here, then, CS termination was not a necessary condition for avoidance conditioning. On the other hand, there are still situations in which it should be recognized that CS termination does provide a sufficient condition for avoidance to occur. In one of the earliest studies on two-factor theory (Brown & Jacobs, 1949), it was shown that CS termination reinforced an instrumental response even though the CS had never before set the occasion for a response that reduced the probability of shock.

Although Herrnstein's analysis forces a reevaluation of two-factor theory, it does not demand that we totally reject the postulated involvement of conditioned fear reactions in avoidance conditioning, for there is too much evidence to the contrary. As Herrnstein pointed out, and as Costello noted, the typical avoidance procedure does contain a Pavlovian conditioning component. When this component is removed, as in the Herrnstein and Hineline (1966) study, we find that it is not a *necessary* condition for avoidance learning. However, other studies, such as that by Brown and Jacobs (1949), do seem to require that CS termination be classified as at least a *sufficient* condition for avoidance.

Even a very substantial revision in our conceptions of avoidance conditioning (which the evidence reviewed by Herrnstein seems to demand) does not require that we reject the proposition that avoidance behaviors and phobias are "similar." It would suggest, however, that the conceptions of the behavior therapists be reevaluated and revised to be more congruent with current laboratory knowledge. The flexibility of the behavior therapies and their openness to changes suggested by experimental evidence is one of their most desirable characteristics in comparison with more analytically based therapies. Analytic conceptions of psychotherapy tend to be more dogmatic and less responsive to laboratory tests. Even though two-factor theory had been shown to be inadequate in some respects, it still may be a more attractive and useful explanation of phobias than some other model that has little or no relationship to laboratory findings. Costello, for example, proposes to view phobias as a "loss of coping reactions," but this alternative provides little more than an interesting description of almost all clinically observed behaviors. In its present form, this model does not provide a description or explanation of phobias per se, nor does it seem open to empirical test.

In conclusion, some few words should be said concerning the possibility that behaviors can be acquired and maintained under different conditions. A behavior acquired under positive reinforcement could, for example, be maintained by negative reinforcement, and vice versa. This same principle could apply in the case of avoidance conditioning where reactions could be acquired according to the principles of two-factor theory but be maintained at a continuing high rate by the principles of Herrnstein's "discrimination hypothesis." Such a proposal would explain why animals in avoidance conditioning seem to become less anxious as learning progresses. In any case, the possibility that phobias are acquired under one set of conditions but maintained under others should be considered further.

REFERENCES

BREGER, L., & MCGAUGH, J. L. Critique and reformulation of "learning theory" approaches to psychotherapy and neurosis. *Psychological Bulletin*, 1965, **63**, 338–358.

BROWN, J. S. Factors affecting self-punitive locomotor behavior. In B. A. Campbell & R. M. Church (Eds.), *Punishment and aversive behavior.* New York: Appleton-Century-Crofts, 1969.

BROWN, J. S., & JACOBS, A. The role of fear in the motivation and acquisition of responses. *Journal of Experimental Psychology*, 1949, **39**, 747–759.

CAMPBELL, B. A., & CHURCH, R. M. (Eds.) *Punishment and aversive behavior.* New York: Appleton-Century-Crofts, 1969.

COSTELLO, C. G. Dissimilarities between conditioned avoidance responses and phobias. *Psychological Review*, 1970, **77**, 250–254.

EYSENCK, H. J., & RACHMAN, S. *The causes and cures of neurosis.* London: Routledge & Kegan Paul, 1965.

HERRNSTEIN, R. J. Method and theory in the study of avoidance. *Psychological Review*, 1969, **76**, 49–69.

HERRNSTEIN, R. J., & HINELINE, P. N. Negative reinforcement as shock-frequency reduction. *Journal of the Experimental Analysis of Behavior*, 1966, **9**, 421–430.

MAIER, S. F., SELIGMAN, M. E. P., & SOLOMON, R. L. Pavlovian fear conditioning and learned helplessness: Effects on escape and avoidance behavior of (a) the CS–UCS contingency and (b) the independence of the US and voluntary responding. In B. A. Campbell & R. M. Church (Eds.), *Punishment and aversive behavior.* New York: Appleton-Century-Crofts, 1969.

MOWRER, O. H. On the dual nature of learning: A re-interpretation of "conditioning" and "problem-solving." *Harvard Educational Review*, 1947, **17**, 102–148.

SIDMAN, M., HERRNSTEIN, R. J., & CONRAD, D. G. Maintenance of avoidance behavior by unavoidable shocks. *Journal of Comparative and Physiological Psychology*, 1957, **50**, 553–557.

SOLOMON, R. L., & WYNNE, L. C. Traumatic avoidance learning: Acquisition in normal dogs. *Psychological Monographs*, 1953, **67**(4, Whole No. 354).

SOLOMON, R. L., & WYNNE, L. C. Traumatic avoidance learning: The principles of anxiety conservation and partial irreversibility. *Psychological Review*, 1954, **61**, 353–385.

WEITZMAN, B. Behavior therapy and psychotherapy. *Psychological Review*, 1967, **74**, 300–317.

(Received October 26, 1970)

1971, Vol. 78, No. 4, 348–351

CONDITIONED AVOIDANCE RESPONSES AND PHOBIAS: A REPLY TO WOLPE AND TO POWELL AND LUMIA[1]

C. G. COSTELLO[2]

University of Calgary

In the paper criticized by Wolpe and by Powell and Lumia, it was argued that conditioned avoidance responses did not provide a plausible experimental paradigm of phobias. These critics have not concerned themselves directly with this argument. Instead, they argue in favor of classically conditioned emotional responses, self-punitive behavior, and learned helplessness as paradigms of phobia acquisition and maintenance. This is of no help in any further valuation of the original proposal.

This paper concerns itself first with Wolpe, then with Powell and Lumia.

Wolpe

In reply to Wolpe (1971), it may be noted that (*a*) I do not entertain the "erroneous belief that behaviorists regard neurosis as an instance of *motor avoidance conditioning*." My paper was concerned with the behavior therapists' theory of *phobia* development. All neurotic conditions are not phobic conditions. It must be admitted that Wolpe in his writings appears to place particular emphasis on the classical conditioning of anxiety responses. On the other hand, as he himself notes, conditioned anxiety is considered to be but one component, albeit the central one, of neurotic habits. In his first book, he lists "physical avoidance of stimuli conditioned to anxiety" as one of "the causal relations of some of the commonest phenomena of human neurotic behavior . . . [Wolpe, 1958, Figure 6, p. 77]." In his latest book, he lists the "avoidance of anxiety evoking stimuli" as one of the "consequences of neurotic anxiety." As the "mechanism" of this consequence, he lists "motor avoidance conditioning [Wolpe, 1969, Table 1, pp. 24–26]." It seems fairly clear that Wolpe accepts the conditioning of avoidance responses as a likely experimental analogue of phobia development. Certainly the weight of the evidence is against any definition of anxiety solely in terms of autonomic responses (for a recent review of some of the relevant literature, see Lang, 1969).

There is no evidence in Eysenck (1968) that he has, as Wolpe claims, adopted classical aversive conditioning theory rather than an avoidance theory for all neurotic conditions. He states clearly that the account of neurotic disorders given in Eysenck and Rachman (1965), and to which I referred in my original paper, "may serve to describe certain types of neurotic disorders . . . [Eysenck, 1968, p. 316]." He does not specify whether these neurotic disorders include phobias. On the other hand, he does not suggest that his theory of incubation should now be adopted to account for phobias. It might be worth noting also that Eysenck chooses a "novel nomenclature" in "putting over" his theory because "the classical account, which is implicitly accepted when we use the classical terms, is somewhat deficient . . . [Eysenck, 1968, pp. 312–313]."

The book by Eysenck and Rachman (1965) was chosen as an example of behaviorists' accounts of phobia development in terms of conditioned avoidance responses. Similar accounts can be found in Ullman and Krasner (1969, pp. 294–298), despite their Skinnerian rather than Hullian emphasis, Meyer and Crisp (1970, pp. 265–271), and Yates (1970, pp. 136–138). Bandura (1969) and Kanfer and Phillips (1970) assume, as I do, that therapists using systematic desensitization in the treatment of phobic disorders accept the conditioned avoidance response analogue, though these authors too are critical of such acceptance.

(*b*) Accounts of phobias in terms of conditioned avoidance appear to be sufficiently prevalent to justify the critique presented in my original paper. If Wolpe wishes to propose an account in purely classical conditioning terms, the plausibility of such an account could very well be examined in a paper similar to my original one. It cannot be done here. The kinds of data with which such an account will have to contend are those of Geer (1968), who

---

[1] The preparation of this paper was supported by Grant 606-7-136 from the Canadian Department of Public Health.

[2] Requests for reprints should be sent to C. G. Costello, Department of Psychology, University of Calgary, Calgary 44, Alberta, Canada.

found that a random association between CS and UCS produced greater emotion defined in terms of galvanic skin responses than either forward or backward classical conditioning.

(*c*) Wolpe agrees with me that Solomon and Wynne's (1954) account of the resistance to extinction of avoidance responses is of doubtful validity. But whereas I suggest this to be so because the two-factor theory on which it is based is of doubtful validity, Wolpe rejects the Solomon and Wynne account because "when 'reality testing' does occur, it is not 'curative.'" But it seems that, on the contrary, there is now considerable evidence that "the avoidance response extinguishes very rapidly following a period of response prevention . . . [Baum, 1970, p. 277]." Furthermore, Wilson and Davison's (1971) review suggests that the provision of counterconditioning agents such as relaxation, drugs, and food probably expedites extinction of avoidance responses in animals by facilitative exposure to the CS. Mathews (1971) makes a similar proposal to account for the success of systematic desensitization in humans.

(*d*) I suggested that conflicts between appetitive and avoidance behaviors were always involved in clinical phobic disorders. Wolpe states that they are not. Unfortunately, neither of us has any experimental evidence to support our views; rather, we rely on our clinical impressions.

(*e*) Seligman's (1970) review of data indicating that not all events are equally associable certainly seems to be relevant to the understanding of Bregman's (1934) failure to produce a conditioned fear response to a neutral CS. Seligman himself points this out. Wolpe may wish to consider some modification of his theory of phobias in the light of these data.

(*f*) Though it may be admitted that conflicting data have appeared on the value of cognitive desensitization and implosive therapy, evidence that therapeutic techniques other than systematic desensitization are effective for the treatment of phobias continues to appear. To those listed in the original paper, we may add (*i*) relaxation training without specific desensitization trials (Kondas, 1967; Paul, 1969; Zeisset, 1968), (*ii*) programmed fantasy (Crowder & Thornton, 1970), (*iii*) leucotomy (Marks, Birley, & Gelder, 1966), (*iv*) antidepressant drugs (Kelly, Guirguis, Frommer, Mitchell-Heggs, & Sargant, 1970), (*v*) expectancy manipulation (Marcia, Rubin, & Efran, 1969), and (*vi*) intravenous acetycholine injections (Sim & Houghton, 1966).

Even if one wishes to agree with Wolpe that "to the extent that methods derived from a causal theory are egregiously successful the theory becomes more plausible . . . ," one should note that not all of the above methods or those listed in the original paper were derived from any one theory of phobia development.

(*g*) Marks (1969, p. 100) reviews the clinical evidence indicating that "patients with phobias often have a shy, dependent, premorbid, personality." On the other hand, there is no experimental evidence on the relationship of dependency needs and phobias. This is one of the reasons why I stated in my original paper that theories alternative to the conditioning theories should be given equal time in the laboratory.

Powell and Lumia

In reply to Powell and Lumia (1971), it may be noted that (*a*) I cannot understand why Powell and Lumia thought I might consider my specific criticisms relating to phobias and conditioned avoidance responses to represent an indictment against behavior therapies in general. I do not so consider them.

(*b*) I agree wholeheartedly that the terms "adaptive" or "adaptiveness" should not be used to describe some inherent characteristic of behavior. That is why I referred only to avoidance responses in the *usual avoidance situation*. Avoidance responses, it seems, may persist under circumstances where they are punished or where free noncontingent shocks are administered. Whether or not avoidance responses under these circumstances provide an adequate experimental analogue for phobias is another matter altogether. Similarly, the disruptive effects of conditioned emotional responses have no direct bearing on the adaptiveness of conditioned avoidance responses in usual or unusual situations. If self-punitive masochisticlike behaviors of the hooded rat or if conditioned emotional responses are to be proposed as the analogues for phobias, the adequacy of the proposals could possibly be evaluated as I evaluated the conditioned avoidance response analogue in my original paper. Again, discussion of the adaptiveness or unadaptiveness of behavior that is being acquired in the noncontingent stage of learned helplessness studies, though perhaps of general interest, has no direct bearing on the adaptiveness or unadaptiveness of conditioned avoidance responses.

Powell and Lumia attempt to save two-factor theory by suggesting that the situation in which avoidance responses are punished contains the element of conflict that I indicated was characteristic of phobias. It is not convincing. I was clearly referring to an approach-avoidance conflict. The self-punitive situations may involve, as Brown (1969) notes, a weak avoidance-avoidance conflict.

(*c*) In presenting evidence against the two-factor theory of avoidance learning, neither Herrnstein's (1969) paper nor my original paper relied solely on experimental data in which CS offset was not a necessary condition for avoidance to occur. Presentation, therefore, by Powell and Lumia of a study (Brown & Jacobs, 1949) showing that CS termination may reinforce an instrumental response has little impact, particularly since Herrnstein (1969) has ably demonstrated that interpretation of such findings is not by any means an easy matter. It should nevertheless be pointed out that there are a number of other studies presenting findings similar to those of Brown and Jacobs (for a recent review see Bitterman & Schoel, 1970).

(*d*) My suggestion that phobias may be the result of "a failure to develop coping behaviors . . ." (not the result of a "loss of coping reactions" as Powell and Lumia wrongly quote me) cannot be developed here but has been further elaborated elsewhere (Costello, 1971). One example of a coping response is the conditioned avoidance response as it is acquired and maintained in the usual avoidance situation. To quote Miller and Weiss (1969):

> Thus, we suggest that having a coping response makes avoidance animals less afraid because relaxation (or safety) becomes conditioned to the response, and then becomes increasingly anticipatory. Perhaps some of the various hypotheses we have discussed may provide the primitive basis for what, at its cognitively more complex human level, we call hope [1969, pp. 360–361].

But then to develop this coping response one has to be a rat, preferably hooded and not albino, and one has to be a lucky enough rat to be assigned to the right condition. Phobic patients appear to be far more helpless than hopeful.

REFERENCES

Bandura, A. *Principles of behavior modification.* New York: Holt, Rinehart & Winston, 1969.

Baum, M. Extinction of avoidance responding through response prevention (flooding). *Psychological Bulletin,* 1970, **74**, 276–284.

Bitterman, M. E., & Schoel, W. M. Instrumental learning in animals: Parameters of reinforcement. *Annual Review of Psychology,* 1970, **21**, 367–436.

Bregman, E. O. An attempt to modify the emotional attitudes of infants by the conditioned response technique. *Journal of Genetic Psychology,* 1934, **45**, 169–198.

Brown, J. S. Factors affecting self-punitive locomotor behavior. In B. A. Campbell & R. M. Church (Eds.), *Punishment and aversive behavior.* New York: Appleton-Century-Crofts, 1969.

Brown, J. S., & Jacobs, A. The role of fear in the motivation and acquisition of responses. *Journal of Experimental Psychology,* 1949, **39**, 747–759.

Costello, C. G. Anxiety and the persisting novelty of input from the autonomic nervous system. *Behavior Therapy,* 1971, in press.

Crowder, J. E., & Thornton, D. W. Effects of systematic desensitization, programmed fantasy and bibliotherapy on a specific fear. *Behaviour Research and Therapy,* 1970, **8**, 35–41.

Eysenck, H. J. A theory of incubation of anxiety/fear responses. *Behaviour Research and Therapy,* 1968, **6**, 309–321.

Eysenck, H. J., & Rachman, S. *The causes and cures of neurosis.* London: Routledge & Kegan Paul, 1965.

Geer, J. H. A test of the classical conditioning model of emotion. *Journal of Personality and Social Psychology,* 1968, **10**, 148–156.

Herrnstein, R. J. Method and theory in the study of avoidance. *Psychological Review,* 1969, **76**, 49–69.

Kanfer, F. H., & Phillips, J. S. *Learning foundations of behavior therapy.* New York: Wiley, 1970.

Kelly, D., Guirguis, W., Frommer, E., Mitchell-Heggs, N., & Sargant, W. Treatment of phobic states with antidepressants. *British Journal of Psychiatry,* 1970, **116**, 387–398.

Kondas, O. Reduction of examination on anxiety and "stage fright" by group desensitization and relaxation. *Behaviour Research and Therapy,* 1967, **5**, 275–281.

Lang, P. J. The mechanics of desensitization and the laboratory study of human fear. In C. M. Frank (Ed.), *Behavior therapy: Appraisal and status.* New York: McGraw-Hill, 1969.

Marcia, J. E., Rubin, B. M., & Efran, J. S. Systematic desensitization: Expectancy change or counterconditioning? *Journal of Abnormal Psychology,* 1969, **74**, 382–387.

Marks, I. M. *Fears and phobias.* New York: Academic Press, 1969.

Marks, I. M., Birley, J. L. T., & Gelder, M. G. Modified leucotomy in severe agoraphobia: A controlled serial inquiry. *British Journal of Psychiatry,* 1966, **112**, 757–869.

Mathews, A. M. Psychophysiological approaches to the investigation of desensitization and related procedures. *Psychological Bulletin,* 1971, **76**, in press.

MEYER, V., & CRISP, A. H. Phobias. In C. G. Costello (Ed.), *Symptoms of psychopathology: A handbook.* New York: Wiley, 1970.

MILLER, N. E., & WEISS, J. M. Effects of the somatic or visceral responses to punishment. In B. A. Campbell & R. M. Church (Eds.), *Punishment and aversive behavior.* New York: Appleton-Century-Crofts, 1969.

PAUL, G. L. Relaxation training and hypnotic suggestion. *Journal of Abnormal Psychology,* 1969, **74**, 425–437.

POWELL, A., & LUMIA, A. R. Avoidance conditioning and behavior therapies: A reply to Costello. *Psychological Review,* 1971, **78**, 344–347.

SELIGMAN, M. E. P. On the generality of the laws of learning. *Psychological Review,* 1970, **77**, 406–418.

SIM, M., & HOUGHTON, H. Phobic anxiety and its treatment. *Journal of Nervous and Mental Disease,* 1966, **143**, 484–491.

SOLOMON, R. L., & WYNNE, L. Traumatic avoidance learning: The principles of anxiety conservation and partial irreversibility. *Psychological Review,* 1954, **61**, 353–385.

ULLMAN, L. P., & KRASNER, L. *A psychological approach to abnormal behavior.* Englewood Cliffs, N. J.: Prentice-Hall, 1969.

WILSON, G. T., & DAVISON, G. C. Processes of fear reduction in systematic desensitization: I. Animal studies. *Psychological Bulletin,* 1971, **76**, 1–14.

WOLPE, J. *Psychotherapy by reciprocal inhibition.* Stanford: Stanford University Press, 1958.

WOLPE, J. *The practice of behavior therapy.* New York: Pergamon Press, 1969.

WOLPE, J. The behavioristic conception of neurosis: A reply to two critics. *Psychological Review,* 1971, **78**, 341–343.

YATES, A. J. *Behavior therapy.* New York: Wiley, 1970.

ZEISSET, R. M. Desensitization and relaxation in the modification of psychiatric patients' interview behavior. *Journal of Abnormal Psychology,* 1968, **73**, 18–24.

(Received February 17, 1971)

Psychological Review
1971, Vol. 78, No. 4, 352–353

A REPLY TO WOLPE

BERNARD WEITZMAN[1]

New School for Social Research, New York

In a reply to Wolpe's criticism of Weitzman's critique of behavior therapy, a number of points are developed. It is asserted that behavior therapists do, in fact, speak of the existence of a "learning theory." It is argued that operational definitions do not, as Wolpe seems to think, supersede theoretical analyses. It is maintained that the issue of the stability of visualizations in desensitization therapy is unresolved. An attempt is made to clarify the basis of apparent misunderstandings of the theoretical arguments presented by Weitzman.

Wolpe (1971) answers criticisms by Weitzman (1967) of the behavioristic view of neurosis. First, he seems to agree that no "learning theory" exists which could be said to be the source of behavior therapy. He suggests that Weitzman is in error in thinking there is an issue here to debate. He accounts for the apparent misunderstanding by referring to "a misleading impression" given by the way in which certain behavior therapists speak. However, in illustrating the position of behavior therapists, Weitzman (1967) quoted Eysenck (1960):

> behavior therapy is an *alternative* type of treatment to psychotherapy . . . it is a *superior* type of treatment, both from the point of view of theoretical background and practical effectiveness . . . in so far as psychotherapy is at all effective, it is so in virtue of certain principles which can be *derived* from learning theory . . . psychotherapy itself, when shorn of its inessential and irrelevant parts, can usefully be considered as a minor part of behavior therapy (p. ix).

The position taken by Eysenck seems plainly and clearly stated and is difficult to "write off" as giving a "misleading impression."

Second, Wolpe argues that "if parallel procedures in cats and men cure the neuroses of both, the species differences must be judged irrelevant in the particular context." Here Wolpe supplies his own context. The context in Weitzman's (1967) discussion was a *theoretical* consideration of the nature of the process through which desensitization is accomplished. In that context, that is, a discussion of the process, species differences have crucial relevance.

Third, Wolpe argues that the criticisms of his use of the term "stimulus" are not valid. In support of this assertion he quotes Lundberg's definition of "stimulus." This definition merely denies the existence of the problem. It does not cogently defend Wolpe's position. Wolpe also defends the validity of his use of the term "stimulus" by pointing to parallel results in desensitization comparing what he calls "extrinsic" versus "imaginal" stimuli. It seems to this writer that the fact that a person may reduce his fear of his boss by confronting his boss and by imagining his boss is not adequate grounds for holding that the confronted boss and the imagined boss are identical stimuli. To argue that stimuli are equivalent which elicit similar responses may have engineering logic but, at this level of analysis, is not satisfying to theoretical reason.

Fourth, Wolpe states that "Patients seem almost invariably able to maintain constant visualization for the standard short spells [5–7 seconds]." He intends, thereby, to undo the force of observations indicating that desensitization may be affected by a process involving the changing of cognitions. Wolpe's report contradicts the experience of the writer which indicates that most patients need at least five seconds to form a clear visualization, although anxiety reactions to the verbal presentation alone are not uncommon, and simply continues a fruitless argument. Since this point is crucial to the conditioning interpretation of desensitization, it would seem to the point to conduct the controlled experimental observations necessary for progress in this debate.

In his discussion of "Weitzman's defense of psychoanalytic psychotherapy," Wolpe makes three points: (*a*) Weitzman does not suggest by what mechanism consonant with psychoanalytic theory desensitization could achieve the lasting disappearance of an "old desire or impulse." The survival of symptomatic behavior was a possibility considered by Freud

---

[1] Requests for reprints should be sent to Bernard Weitzman, Graduate Faculty, New School for Social Research, 66 West 12th Street, New York, New York 10011.

(1936). The present writer made no such claim for desensitization. The argument presented may be summarized as follows: If treatment by desensitization does not entail symptom substitution, then we must consider, as Freud did, the possibility that we were treating a habit not maintained by "old desires and impulses." In other words, it was being pointed out that psychoanalytic thought could entertain a "habit" conception of certain symptoms.

(*b*) Wolpe understands Weitzman as proposing that "desensitization may be a repression-lifting mechanism." In this section of the paper, Weitzman is, in fact, exploring the opposite possibility, that is, that desensitization may be a repression-augmenting mechanism. That is, the possibility is presented that any ego-syntonic alteration of object relations, by any therapeutic means, may *increase* the repressive potential of the ego. In another section of the paper, Weitzman describes an application of the desensitization technique to the analysis of dreams. This use of desensitization is described as facilitating the overcoming of resistance and might, thus, accurately be characterized as "repression lifting." Behavior therapists, however, are not typically in a position to replicate this observation.

(*c*) Finally, Wolpe appears to have missed the main point of the analysis of the symbolic function of the symptom. It was argued there that the term "symptom treatment" is a misnomer. It implies a compartmentalization of man which is antagonistic to psychoanalytic thought. In fact, from the point of view of psychoanalysis, there can be no "symptom treatment." This assertion is based on the postulated symbolic relationship between symptom and dynamic cause which is basic to psychoanalytic thought. Once this relationship is grasped, it becomes clear that in treating a symptom one is, of necessity, affecting the status of the "internal danger."

REFERENCES

EYSENCK, H. J. (Ed.), *Behavior therapy and the neurosis*. New York: Pergamon Press, 1960.

FREUD, S. *The problem of anxiety*. New York: Norton, 1936.

WEITZMAN, B. Behavior therapy and psychotherapy. *Psychological Review*, 1967, **74**, 300-317.

WOLPE, J. The behavioristic conception of neurosis: A reply to two critics. *Psychological Review*, 1971, **78**, 341-343.

(Received March 16, 1971)

Psychological Review
1971, Vol. 78, No. 4, 354

LETTERS TO THE EDITOR

A NONMEDIATIONAL IDEOMOTOR THEORY

Greenwald's ideomotor theory (Greenwald, 1970) makes two claims on page 93, the first of which is incorrectly interpreted, the second of which is logically questionable, if not impossible.

He postulates that "ideomotor analysis assumes that responses are represented in the form of images of their consequences . . . [p. 93]." This point is substantiated by spectrographic analyses of speech production, in that the initial /s/ in /stru/ shows the contextual influence of the vowel which follows three phonemes later (Halwes & Jenkins, 1971). However, as Halwes and Jenkins demonstrate, the "machinery" that is required to generate such a context effect must be more powerful than merely "subgoals . . . as images of sensory feedback . . . [Greenwald, 1970, p. 93]."

In discussing symbolic mediators, Greenwald claims that "'ball' is assumed to abstract the common properties of a variety of stimuli . . . [p. 93]." However, the number of features *common* to all instances of a conceptual mediator, such as "ball," rapidly approaches an infinitesimal, if not null, set. And to assert that a ball is any type of an object that is set in motion during a game will place the mediator into the category of "ball-relevant responses," which merely begs the question. Greenwald's claim would have been more accurate had he used the verb "subtract" instead of "abstract." But a truly abstractive theory of concepts would be *generative* (deriving a situation-specific response from rules operating on abstract entities), and *not* mediational (pairing a specific abstracted [sic] feature with a specific response).

REFERENCES

GREENWALD, A. G. Sensory feedback mechanisms in performance control: With special reference to the ideo-motor mechanism. *Psychological Review*, 1970, **77**, 73–99.

HALWES, T., & JENKINS, J. J. Problem of serial order in behavior is not resolved by context-sensitive associative memory models. *Psychological Review*, 1971, **78**, 122–129.

(Received February 26, 1971)

Frederick H. Steinheiser, Jr.
Center for Research in Human Learning
University of Minnesota
Minneapolis, Minnesota 55455

SPENCE'S PREDICTION ABOUT REVERSAL-SHIFT BEHAVIOR

A simulation (Wolford & Bower, 1969) of Spence's original discrimination-learning model (Spence, 1936) demonstrated that in some cases subjects should execute a reversal shift more rapidly than an extradimensional (nonreversal) shift. This finding was thought to be inconsistent with interpretations offered by mediational theorists (e.g., Kendler & Kendler, 1962) that Spence's single-unit formulation predicted that a reversal should always be slower than an extradimensional shift. The latter interpretation is correct, however, where reversal and extradimensional shifts are compared for a discrimination-shift study involving a counterbalanced design with two orthogonal dimensions (e.g., brightness and shape), each of which contain two values (e.g., black, white, triangle, circle). Regardless of what start vectors were used, two independent simulations, by two of the correspondents (MAH and GW), found that Spence's model always predicted a faster extradimensional than reversal shift when preshift training was carried to a criterion of 10 successive correct responses. In sum, Spence's theory, depending on the initial start vectors, can predict faster reversal or extradimensional shifts. But when it comes to counterbalanced designs, with no overtraining, the model is unequivocal; reversals should be slower than extradimensional shifts.

REFERENCES

KENDLER, H. H., & KENDLER, T. S. Vertical and horizontal processes in problem solving. *Psychological Review*, 1962, **69**, 1–16.

SPENCE, K. W. The nature of discrimination learning in animals. *Psychological Review*, 1936, **43**, 427–449.

WOLFORD, G., & BOWER, G. H. Continuity theory revisited: Rejected for the wrong reasons? *Psychological Review*, 1969, **76**, 515–518.

(Received April 1, 1971)

Howard H. Kendler,
Morton A. Hirschberg
Department of Psychology
University of California, Santa Barbara
Santa Barbara, California 93106
and
George Wolford
Department of Psychology
Dartmouth College
Hanover, New Hampshire 03755

The British Journal of Psychology

Volume 62 Part 2 May 1971

R. CONRAD The effect of vocalizing on comprehension in the profoundly deaf

GILLIAN COHEN Differential effects of irrelevant dimensions in three shape recognition tasks

N. J. WADE AND H. SCHÖNE The influence of force magnitude on the perception of body position. I. Effect of head posture

JAMES T. REASON AND ASHTON GRAYBIEL The effect of varying the time interval between equal and opposite Coriolis accelerations

D. M. A. GRONWALL AND H. SAMPSON Ocular dominance: a test of two hypotheses

C. D. FRITH Strategies in rotary pursuit tracking

H. J. EYSENCK AND J. E. GRAY Reminiscence and the shape of the learning curve as a function of subjects' ability level on the pursuit rotor

J. M. HINTZ AND T. M. NELSON Haptic aesthetic value of the golden section

KATE LOEWENTHAL A study of imperfectly acquired vocabulary

PETER W. SHEEHAN The role of imagery in incidental learning

DAVID A. BLIZARD Situational determinants of open-field behaviour in *Mus musculus*

A. J. SUTTON The use of quadratic discriminant analysis for the measurement of profile distance in social perception

£2.50 (US $8.50) each, plus postage. Annual subscription for four issues £7.50 (US $25.00)

CAMBRIDGE UNIVERSITY PRESS

Bentley House, 200 Euston Road, London NW1 2DB
American Branch: 32 East 57th Street, New York, NY 10022

now available in reprint

PSYCHOLOGICAL MONOGRAPHS:
GENERAL AND APPLIED

(American Psychological Association)
Vols. 1-52. Princeton; etc. 1895-1940

Psychological Monographs makes available longer researches, treatises, or collections of laboratory studies which it was important to publish in their entirety and as they were received. The length of these studies usually excluded them from other journals, but the depth and detail of their findings make them essential for advanced study in psychology.

Publication of the *Monographs* began in 1895 as a collateral activity of the *Psychological Review* and continued as a major source of longer, more comprehensive articles until 1967, when the individual journals published by the American Psychological Association began to publish supplementary *Monographs* in their own area.

For a complete list of available titles, and their prices,
please write to our Sales Department.

JOHNSON REPRINT CORPORATION
111 Fifth Avenue, New York, N.Y. 10003

R.C. 19.7.71

KC
Asked 2/
issue to Tu
Librarian
R Das
26/10/71

PSYCHOLOGICAL REVIEW

GEORGE MANDLER, EDITOR
JEAN M. MANDLER, ASSOCIATE EDITOR
UNIVERSITY OF CALIFORNIA, SAN DIEGO

Vol. 78, No. 3 **May 1971**

Articles

Integration Theory and Attitude Change **Norman H. Anderson** 171

A Model of the Perception of Simple Geometric Figures **Paul C. Vitz and Thomas C. Todd** 207

The Dynamics of Structured Personality Tests: 1971 **Douglas N. Jackson** 229

Discrepancy Hypotheses: Methodological and Theoretical Considerations................**Hoben Thomas** 249

Theoretical Notes

Comments on Biederman's "Continuity Theory Revisited: A Failure in a Basic Assumption" **Daniel B. Berch** 260

Hypothesis Analysis of Conjunctive Concept-Learning Situations......................**Lance A. Miller** 262

Comment on "General Shape Function Model of Learning with Applications in Psychobiology" **Joel P. Wiesen** 272

Letters to the Editor 274

Published Bimonthly
by The American Psychological Association, Inc.

Editorial Consultants

FRED ATTNEAVE
DARYL BEM
DANIEL E. BERLYNE
DONN BYRNE
HERBERT H. CLARK
BARRY COLLINS
STEWART COOK
ROBYN DAWES
A. J. DEUTSCH
MARTIN FISHBEIN
LEO GANZ
WENDELL GARNER
JAMES GREENO
LEE W. GREGG
G. ROBERT GRICE
JULIAN HOCHBERG
PAUL HOLLAND
W. K. HONIG
JANELLEN HUTTENLOCHER
H. M. JENKINS
MARVIN LEVINE
PERRY LONDON
SAMUEL MESSICK
NORMAN MILLER
D. A. NORMAN
DAVID PREMACK
FRANK RESTLE
RICHARD ROSE
PERCY TANNENBAUM
H. S. TERRACE
JOHN THEIOS
JERRY WIGGINS
H. A. WITKIN

(A large group of psychologists has kindly agreed to assist in the editing of manuscripts submitted to the *Psychological Review*. In each issue the Editor will list in the *Review* those editorial consultants who have helped him process manuscripts for that issue, so that their efforts may be gratefully acknowledged.)

The *Psychological Review* is devoted to articles of theoretical significance to any area of scientific endeavor in psychology. Except for occasional articles solicited by the Editor, manuscripts exceeding 7,500 words will not be accepted. Ordinarily, manuscripts which consist primarily of original reports of research or of reviews of the literature should be submitted to other journals. Methodological and statistical articles are not appropriate for the *Psychological Review*. See the editorials in the *Review* (1965, 72, 1–2; 1971, 78, 1–2) for further details.

Beginning with the 1971 issue, Letters to the Editor will be published on an experimental basis. Letters not exceeding 250 words on topics of theoretical interest should be submitted in duplicate.

Manuscripts must be accompanied by an abstract of 100 to 120 words typed on a separate sheet of paper. The abstract should conform to the style of *Psychological Abstracts*. Detailed instructions for preparation of the abstracts appeared in the *American Psychologist* (1961, 16, 833), or may be obtained from the APA Central Office.

All manuscripts must be submitted in triplicate. In the reference list, give journal titles in full; do not abbreviate. Figures should be submitted as glossy prints; duplicate figures may be photographic or pencil-drawn copies. Authors should retain a copy of the manuscript in case there is loss in the mail.

Because of the large number of manuscripts submitted, there is an inevitable publication lag of several months. Authors may avoid this delay if they are prepared to pay the costs of publishing their own articles; the appearance of articles by other contributors is not thereby delayed.

Manuscripts should be addressed to the Editor, George Mandler, Department of Psychology, University of California, San Diego, La Jolla, California 92037.

Copyrighted material. The following policies govern reprinting or other forms of reproduction of materials copyrighted by the American Psychological Association: Written permission must be obtained for copying or reprinting any tables, figures, or text of more than 500 words in length. Persons or organizations who obtain permission to reproduce any such copyrighted materials must include on the first page of the reproduced materials the exact copyright notice which appeared on the Association's original publication. There will be a charge of $10.00 per page for such permission except when the requester is the author of the article to be reprinted, or when the article is to be reproduced in a limited number of copies solely for instructional purposes.

HELEN ORR
Managing Editor

LORRAINE BIROS
Technical Editor

HENRY S. SALMON
Advertising and Business Manager

ROBERT D. GORMAN
Subscription Manager

PUBLISHED BIMONTHLY BY THE
AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
Prince and Lemon Streets, Lancaster, Pa.
1200 Seventeenth St., N. W., Washington, D. C. 20036

$10.00 volume

$2.00 issue

Second-class postage paid at Lancaster, Pa.

Send all communications, including address changes, to 1200 Seventeenth St., N. W., Washington, D. C. 20036. Address changes must arrive by the tenth of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

405
4·11·71
Lelo

PSYCHOLOGICAL
REVIEW

VOL. 78, NO. 3 MAY 1971

INTEGRATION THEORY AND ATTITUDE CHANGE[1]

NORMAN H. ANDERSON[2]
University of California, San Diego

A theory of information integration is applied to attitudes and social judgments, based on a principle of information integration. For quantitative analysis, a simple but general algebraic model of judgment is used, in which each informational stimulus is characterized by two parameters, scale value and weight. Functional measurement procedures are employed to derive equal-interval scales of parameter values. Exact tests based on analysis of variance are given for four applications of the model, and these applications are reconsidered under the further restriction imposed by the averaging hypothesis. Qualitative comparisons are made to several other theories of attitude change. Tendencies toward balance and congruity are shown to be consequences of the principle of information integration. Critical tests between integration theory, and balance and congruity theories are also suggested. Similar comparisons are made to summation theory, logical-consistency theory, assimilation-contrast theory, and similarity-attraction theory. Molar and molecular analyses of communication structure are considered briefly and the analysis of inconsistency resolution within integration theory is also discussed. Finally, it is noted that integration theory has had reasonable success in the areas of learning, perception, judgment, decision making, and personality impressions, as well as attitude change. It may thus provide a beginning to a unified general theory.

Attitude change stands out from most areas of experimental psychology in the nature of its stimuli. In even the simplest investigations of attitudes and opinions, the stimuli typically carry information at a cognitive level not often reached in other areas of research. Informational stimuli continually impinge on the person, in life or in the laboratory, and he must integrate them with one another as well as with his prior opinions and attitudes. Social judgments are typically based on a cumulation of various pieces of information, sometimes of the most diverse nature. Factual and hearsay evidence, rumors, prestige associations, gesture and appearance, may all bear on the final attitude. Information integration is thus fundamental in attitude change.

---

[1] This work was supported by grants from the National Science Foundation (GB-6666) and the Office of Education (Project No. 9-0399) to the author, and from the National Institute of Mental Health to the Center for Human Information Processing, University of California, San Diego. This paper owes much to many people. I wish to express my enduring appreciation to David A. Grant and Carl I. Hovland who stimulated my interest in the method and content of this work. I owe special thanks to Sam Himmelfarb and Rhoda Lindner for detailed, insightful criticism, and to Robyn Dawes and Robert Wyer for many helpful comments. I am also indebted for comments, not always heeded, on particular points by Donn Byrne, Don Campbell, Gerald Clore, Rudolf Cohen, Harry Gollob, Philip Gough, Kenneth Hammond, Reid Hastie, Clyde Hendrick, Chester Insko, Martin Kaplan, Jacques Kaswan, Howard Leventhal, Theodore Newcomb, Ramon Rhine, Milton Rosenberg, Ralph Rosnow, Marshall Segall, Jim Shanteau, Marvin Shaw, Paul Slovic, Raymond Smith, and Warner Wilson.

[2] Requests for reprints should be sent to Norman H. Anderson, Department of Psychology, University of California, San Diego, La Jolla, California 92037.

© 1971 by the American Psychological Association, Inc.

In previous research, the writer has been working on a general theory of information integration. Experimental applications have ranged from personality impression formation to decision making and psychophysical judgment. The present paper gives a more detailed exposition of integration theory in attitude change, developing further the theme of some of the first reports in this research program (Anderson, 1959a; Anderson & Hovland, 1957) and of more recent work (Anderson, 1968a).

INTEGRATION THEORY

Integration Model

A basic distinction in the present formulation is that between weight and value. Each piece of information is represented by two parameters: a scale value, s, and a weight, w. The value is the location of the informational stimulus along the dimension of judgment. The weight represents the psychological importance of the information. It is important to note that both s and w will depend on the dimension of judgment as well as the individual. For simplicity, only a single dimension of attitude or judgment will be explicitly considered. The same piece of information may have quite different value and importance on different dimensions, or for different individuals on the same dimension.

The theoretical model used in most of this work can be written as a weighted sum:

$$R = C + \sum_{i=0} w_i s_i. \qquad [1]$$

Usually, R is taken to be the overt response, measured on a numerical scale. In some cases, however, R might be considered as the attitude underlying an overt yes–no decision. The summation in Equation 1 is over all relevant informational stimuli. The contribution of Stimulus i is just its weight w_i times its value s_i. The constant, C, allows for an arbitrary zero in the response scale and will not be explicitly considered here. The first term in the sum, $w_0 s_0$, represents the initial opinion, prior to receiving the informational stimuli.

Two basic algebraic operations, adding and multiplying, are involved in the integration model. In simplest form, the total effect of any one informational stimulus is the product of its weight and its scale value. The total effect of two or more informational stimuli is the weighted sum of their scale values. These two algebraic operations correspond to two classes of experimental manipulations in attitude experiments as noted below.

Despite its simple nature, the algebraic judgment model has considerable flexibility. It includes adding, averaging, subtracting, dividing, and multiplying models, as well as proportional-change and linear operator models. Special cases have been studied by numerous investigators in varied areas as cited here and elsewhere (Anderson, 1970a; S. Rosenberg, 1968).

Empirical Support

The writer's program of research in information integration has been based on a systematic exploitation of the algebraic judgment model. A substantial portion of the work has dealt with personality impressions (see Anderson, 1968a), including the study of stimulus interaction and configural use of cues (Anderson, 1965b; Anderson & Jacobson, 1965; Lampel & Anderson, 1968; Sidowski & Anderson, 1967). Other applications include decision making (Anderson & Shanteau, 1970; Shanteau, 1970a), psychophysical judgment (Parducci, Thaler, & Anderson, 1968; Weiss & Anderson, 1969), illusions (Anderson, 1970b; Massaro & Anderson, 1971), learning (Anderson, 1969b; Friedman, Carterette, & Anderson, 1968; Himmelfarb, 1970), and attitude change (Sawyers & Anderson, 1971).

The integration model has done well, both qualitatively and quantitatively. It rests upon a substantial data base. Although a major part of the supporting experimentation has been done with personality impressions, the same principles may be expected to apply to the more traditional communication-persuasion paradigm. Indeed, personality impression formation involves attitude change toward a particular individual and can be viewed as a miniature attitude change situation. Each trait adjective corresponds to a communication whose weight can be controlled by attributing it to various sources, for instance, or by manipulating its reliability. The development of personality impressions and attitudes both involve the integration of information into evaluative judgments that have social relevance. From the present standpoint, an a priori distinction between them would be artificial.

Valuation and Integration

The two fundamental operations in integration theory are valuation and integration. Valuation, in model terms, involves the determinants and the measurement of the w and s parameters. Integration comprises the ways in which the several stimuli are combined. The judgment task may affect the integration rule and certainly will affect the valuation process. In general, however, valuation and integration may conveniently and effectively be kept separate as long as the stimuli do not interact.

When the stimuli do interact, valuation and integration will become interlinked. As a working rule, it will be assumed here that stimulus interaction primarily affects the valuation process, and that the changed parameter values are then employed in the integration.

Most of the experimental work to date has used the assumption that the w and s values for an informational stimulus remain constant, regardless of what other stimuli it may be combined with. This is convenient in tests of the model, particularly in the early stages of development. Fortunately, the assumption that the stimuli do not interact seems to be justified in an important group of situations.

More interesting though more difficult problems arise in the many situations in which context-induced changes may occur. Primacy effects (Anderson, 1965b; Asch, 1946), positive context effects (Anderson, 1966a; Kaplan, 1971), discounting effects (Anderson & Jacobson, 1965; Lampel & Anderson, 1968; Schümer & Cohen, 1968), and differential weighting (Oden & Anderson, 1971) have all been studied in personality impressions with some theoretical success. Such problems of configural cue usage are even more important in the general study of attitude change as the latter discussion of the source-communication relation will show.

The problem of stimulus configurality also bears directly on the problem of cognitive structure. A personality impression, for example, or even a trait adjective, obviously has a more complex cognitive structure than can be represented on a single dimension. A multidimensional representation would do better, of course, and the present development may be considered as applying to each of several dimensions in turn.

However, the undoubted success of the dimensional view in psychology can lead to preoccupation with a restricted set of problems, and to a possibly artificial view of cognitive structure. Information integration, for example, may itself occur at a more basic level than the judgment, and the latter may be merely the valuation of the structure along the specified task dimension (Anderson, 1967). The standard dimensional view will be used in this paper, but its probable limitations need particular attention in the study of stimulus interaction. For example, the data suggest that affective and semantic inconsistency produce about equal discounting (Anderson & Jacobson, 1965). In inconsistent combinations, the response must be based on the stimulus configuration, and it is problematic how far standard dimensional representations of the stimulus will go.

Application to Attitude Change

In attitude change, the source-communication-issue schema developed most extensively by Hovland and his associates (e.g., Hovland, 1957; Hovland, Janis, & Kelley, 1953) provides a useful way to organize many problems. A communication from some source provides information that is relevant to the person's attitude or opinion on some issue. The effective stimulus is then the source-communication combination. The present use of two parameters, weight and scale value, to represent the source-communication parallels the approach of Anderson and Hovland (1957).

The scale value of the source-communication corresponds to its position on the attitude dimension. For many purposes, this can be taken as the position advocated by the source-communication. This value will be primarily determined by the semantic properties of the communication, as they are interpreted within the person's value system. Conceptually, it is sometimes useful to consider the scale value as the attitude that would be left unchanged by the communication.

The weight parameter is more important theoretically than the scale value because of the many determinants of the weight. A primary determinant of w_i will be characteristics of the source. Thus, source status, reliability, and expertise will all affect w_i. Moreover, w_0, the weight parameter of the initial opinion, will reflect personality characteristics, such as persuasibility, as well as ego involvement and strength of prior opinion. In addition, the evidence indicates that primacy and recency effects are to be interpreted in terms of the weight parameter as discussed later.

Because of its close relationship with functional measurement, integration theory emphasizes a more quantitative approach to communication-persuasion processes than has been usual in attitude theory. Traditional approaches have been more concerned with qualitative analysis of valuation than with integration per se. Since functional measurement opens up possibilities for direct measurement of the stimuli, the present approach is less exclusively concerned with qualitative analysis. Moreover, because the theory rests on a principle of information integration, the experimental program has emphasized the study of information variables in the stimulus. As will become apparent, some of the traditional problems of attitude theory are particularly amenable to analysis within a conceptual framework based on information integration.

Functional Measurement

The power of the algebraic judgment model has depended on the concomitant development of a theory of functional measurement. This scaling theory makes it possible to get the equal-interval scales that are needed in the quantitative tests of the model. The basic ideas of functional measurement were given in two early papers (Anderson, 1962a, 1962b), and a more systematic exposition in psychophysical judgment is given in Anderson (1970a). A brief sketch will be given here; the later theoretical applications will exhibit functional measurement in action.

A central idea of functional measurement is that measurement scales are derivative from substantive theory. Measurement thus involves three simultaneous problems (Anderson, 1970a, Figure 1): (*a*) Problem 1—to measure the subjective stimulus values on equal-interval scales; (*b*) Problem 2—to measure the subjective response value on an equal-interval scale; and (*c*) Problem 3—to find the psychological law that relates the subjective values of stimuli and response. All three problems are to be solved together. That this can sometimes be done quite simply is shown in previous experimental work using functional measurement procedures (e.g., Anderson, 1962a). These procedures are illustrated in the next section.

Nearly all the experimental work in this research program has used numerical response measures such as ratings. But rating scales may be only ordinal, while the exact tests of the model require an equal-interval scale. Functional measurement allows transformation of an ordinal scale into an equal-interval scale on the response side. Functional scales on the stimulus side can then be obtained. The algebraic judgment model is used as the scaling frame so that the validity of the scales depends on the validity of the model. Substantive theory and measurement theory are thus cofunctional in development.

Functional measurement may be contrasted with Thurstonian scaling in several respects, two of which should be noted here. First, the Thurstonian approach is largely concerned with stimulus scaling, which is traditionally seen as a preliminary to the substantive inquiry. Such an approach has its attractions, but it bases the substantive investigation on a separate scaling theory of uncertain validity and relevance. Many workers, as a consequence, have attempted to develop "measurement free" methods that use only rank-order information. The present approach, in contrast, treats scaling and theory as inherently interlinked at every level of consideration.

A second difference between functional measurement and Thurstonian scaling is in the treatment of individual differences, a problem of some importance in attitudes and opinions. As is well known, Thurstonian paired-comparison scales of attitudinal stimuli almost necessarily rely on pooled group data. Most people will have clear preferences, and the resultant 0–1 probabilities disallow scaling at the individual level. Group scales (Bock & Jones, 1968, pp. 1–2) have normative and sociological value, as in Coomb's (1967) replication of Thurstone's (1927) scale of seriousness of offenses. However, their role in the study of psychological processes is severely limited. Functional measurement has the important advantage that it can operate directly at the level of the individual.

Conceptually, the response scaling feature (Anderson, 1962b, 1970a; Bogartz & Wackwitz, 1970, 1971) is essential to functional measurement. The logic of this approach rests on "using the postulated behavior laws to induce a scaling on the dependent variable" (Anderson, 1962b). Indeed, it would not be possible to consider this approach as a general theory of measurement without systematic methods for transforming ordinal response measures to an interval scale. Experimental illustrations are given in Anderson and Jacobson (1968) and Weiss and Anderson (in press).

For the kinds of situations considered in this article, however, there is increasing evidence that ordinary rating procedures can provide interval scales simply and directly, at least with reasonable experimental precautions (e.g., Anderson, 1967; Himmelfarb & Senn, 1969; Shanteau & Anderson, 1969). In this article, therefore, it will be assumed that the observed response is on an interval scale. This assumption is directly testable since an inadequate scale would tend to disconfirm the model predictions. Conversely, if the model does fit the data, response scale and model are jointly validated.

From the experimental standpoint, an important methodological feature of functional measurement is the use of stimulus combinatorics. Each informational stimulus is systematically placed in several different combinations. This experimental strategy embodies a direct approach to the substantive problem of stimulus integration. Happily, it also allows the effect of the given stimulus to be factored out and scaled. Extensive use is made of factorial designs, therefore, with analysis of variance a principal tool.

Conjoint measurement (Krantz & Tversky, 1970; Luce & Tukey, 1964) has certain similarities to functional measurement, but as a matter of principle restricts itself to rank-order data. Its principal virtue seems to be that it can provide rank-order tests of goodness of fit for certain models. That virtue bypasses the response scaling problem, which could certainly be advantageous, though practical difficulties seem to have limited the use of these techniques (Tukey, 1969; Zinnes, 1969). Functional measurement, in contrast, enables one to use whatever metric information is in the overt response. In that sense, the difference between functional and conjoint measurement parallels that between parametric and nonparametric tests (Anderson, 1961a). At a more fundamental level, the functional measurement approach makes measurement theory an integral part of substantive psychological theory. As one consequence, it has the flexibility needed to cope with nonadditivities caused by averaging processes and certain kinds of stimulus interactions.

MATHEMATICAL APPLICATIONS

This section details several applications of the integration model, and much of it can be skipped initially. After the next two subsections, therefore, and the initial remarks on the averaging hypothesis, it may be desirable to go to the main section of the paper, Relations to Other Theories.

Basic Integration Rules

The applications below all rest on two kinds of experimental manipulations: of the weight parameter and of the scale value. These correspond to three basic integration rules. The first is a multiplying rule which specifies the total effect of a communication to be the product of its weight and its value. The rationale for this multiplying rule can be illustrated by considering any method of manipulating weight. Source credibility, for example, does not of itself produce opinion change, but only affects degree of acceptance of the communication. A decrease in source credibility would thus act by attenuating the effect of the communication.

The second basic integration rule specifies that two communications will follow some kind of adding rule, as in Application 1 below. An adding-type rule is intuitively attractive and has been employed in some form in almost all attempts at quantitative analysis.

However, there is considerable evidence for an averaging rule instead of an adding rule. Mathematically, the two rules have considerable similarity, though the averaging rule tends to be more difficult to work with. In many cases, the two rules make the same prediction, but they disagree markedly in some important situations as discussed under Averaging Theory.

Psychologically, adding and averaging are quite different. In a strict adding model, the informational stimuli can be completely independent in their action. In contrast, averaging always implies some degree of cognitive interaction because the weight of any one informational stimulus necessarily depends on all the others. Accordingly, averaging is considered as a third basic rule and considered separately in the next section.

A comment on terminology may help avoid some ambiguity in the use of the terms "adding" and "averaging." Because they make the same predictions under certain conditions, there may be neither need nor basis to make a distinction. In many of the published papers that have employed one or the other model, no distinction has been necessary. In line with mathematical usage, therefore, the term adding will be used generically to include both strict adding and averaging models. When the discussion is to be restricted to one or the other, this will be indicated. The results of the following four applications do not all hold for averaging models as discussed in the next main section on averaging theory.

Application 1: Variation in s

This application is appropriate when two (or more) communications that have different scale values on a given issue are combined. In the experiment cited below, for example, the communications were paragraphs that described good and bad actions of United States Presidents.

It is assumed that the communications are presented according to a Row × Column factorial design as illustrated in Table 1. The row and column factors each correspond to one source, and the levels within each factor correspond to communications attributed to the row source and column source, respectively. The source may be only implicit, when the communications are presented as historical fact, for example.

Under an adding model, the theoretical entry in each cell of the design is

$$R_{ij} = w_0 s_0 + w_{\mathrm{R}} s_{\mathrm{R}i} + w_{\mathrm{C}} s_{\mathrm{C}j}. \qquad [2]$$

Here w_{R} and w_{C} are the communication weights associated with the row and column sources; $s_{\mathrm{R}i}$ is the scale value of the communication coming from Source R in row i; $s_{\mathrm{C}j}$ is the scale value of the communication from Source C in column j. Each opinion is thus considered as the resultant of three informational stimuli: the two communications and the initial opinion. The initial opinion will be assumed equal to zero, a simplification that does not affect the conclusions.

Equation 2 imposes two assumptions: The first is that the communications do not interact or change scale value when combined. The second is that the row weight, w_{R}, is constant over rows and analogously for the column weight. Since w_{R} represents the joint effect of the row source and the row communication, this virtually requires that each row communication have the same natural weight and that source and communication do not interact.

These restrictive assumptions should be kept in mind when designing an experiment. They are, however, more restrictive than necessary for an adding model. Application 4, below, allows the weights to vary across rows and/or columns and still yields exact predictions.

Test of fit. For this design, the model makes a very simple prediction: each two rows differ by a constant. For example, the difference between rows 1 and 2 in each column of Table 1 is $w_{\mathrm{R}}(s_{\mathrm{R}1} - s_{\mathrm{R}2})$.

If the model is correct, then the raw data should show a very simple pattern. Graphically, the three rows of data should plot as three parallel curves except for sampling error. An example based on Table 1 is given in Figure 1 in the next section. Statistically, this means that the Row × Column interaction is zero in principle and nonsignificant in practice. Thus, regular analysis of variance provides a simple and powerful test of goodness of fit.

It may be reemphasized that this parallelism test can be made directly on the raw data. The person receives the two informational stimuli and then makes his response on a numerical scale. The test is made directly on these responses. Separate esti-

TABLE 1

INTEGRATION MODEL ILLUSTRATED FOR TWO-COMMUNICATION FACTORIAL DESIGN WITH VARIOUS COMMUNICATIONS ATTRIBUTED TO EACH OF TWO SOURCES

A. Formal model expressions

Source R (w_R)	Source C (w_C)				
	Communication C1	Communication C2	Communication C3	Communication C4	M
Communication R1	$w_R s_{R1} + w_C s_{C1}$	$w_R s_{R1} + w_C s_{C2}$	$w_R s_{R1} + w_C s_{C3}$	$w_R s_{R1} + w_C s_{C4}$	$w_R s_{R1} + w_C \bar{s}_C$
Communication R2	$w_R s_{R2} + w_C s_{C1}$	$w_R s_{R2} + w_C s_{C2}$	$w_R s_{R2} + w_C s_{C3}$	$w_R s_{R2} + w_C s_{C4}$	$w_R s_{R2} + w_C \bar{s}_C$
Communication R3	$w_R s_{R3} + w_C s_{C1}$	$w_R s_{R3} + w_C s_{C2}$	$w_R s_{R3} + w_C s_{C3}$	$w_R s_{R3} + w_C s_{C4}$	$w_R s_{R3} + w_C \bar{s}_C$
M	$w_R \bar{s}_R + w_C s_{C1}$	$w_R \bar{s}_R + w_C s_{C2}$	$w_R \bar{s}_R + w_C s_{C3}$	$w_R \bar{s}_R + w_C s_{C4}$	

B. Numerical example

Source R ($w_R = 1$)	Source C ($w_C = 2$)				
	$s_{C1} = 1$	$s_{C2} = 2$	$s_{C3} = 3$	$s_{C4} = 6$	M
$s_{R1} = 1$	3	5	7	13	7
$s_{R2} = 2$	4	6	8	14	8
$s_{R3} = 6$	8	10	12	18	12
M	5	7	9	15	

mates of the scale values of the communications are not required. This is invaluable in attitude research because prior scaling of the communications is often undesirable or infeasible.

Application 1 has special interest because it holds when the communications are presented in strictly serial order as is typical of attitude change situations. The row factor then corresponds to the first communication, the column factor to the second communication, etc. Applications of such serial-factor designs have been made in a number of other areas (e.g., Anderson, 1964d; Shanteau, 1970a; Weiss & Anderson, 1969). Under certain conditions, it is possible to construct the serial-position curve knowing only the response at the end of the sequence.

Functional scales. A striking property of the integration model is that it provides a method for scaling the communications. In Table 1, the column means are linear functions of the s_{Cj},

$$\bar{R}_j = w_R \bar{s}_R + w_C s_{Cj}, \qquad [3]$$

where the mean row value, $\bar{s}_R$, is a constant. The observed column means, therefore, are estimates of the communication values on an equal-interval scale.

This is illustrated in the numerical example in the lower half of Table 1. The column means are twice the column scale values plus 3; the row means are the row scale values plus 6. These interval scales have arbitrary zero and unit and are unique only up to a linear transformation. This needs to be kept in mind for scaling purposes; more than two row stimuli would ordinarily be needed to estimate any scale values for the row stimuli. Of course, no more than two rows are needed for the test of fit, and two rows suffice for scaling under certain conditions when the same stimuli are used in each factor of the design (e.g., Weiss & Anderson, 1969).

These results are not restricted to two-way designs but hold for any number of factors. The first experimental application of functional measurement (Anderson, 1962a) used a three-way design for personality impressions.

Experimental example. Figure 1 shows an experimental test of integration theory. Subjects judged general statesmanship of various United States Presidents, each de-

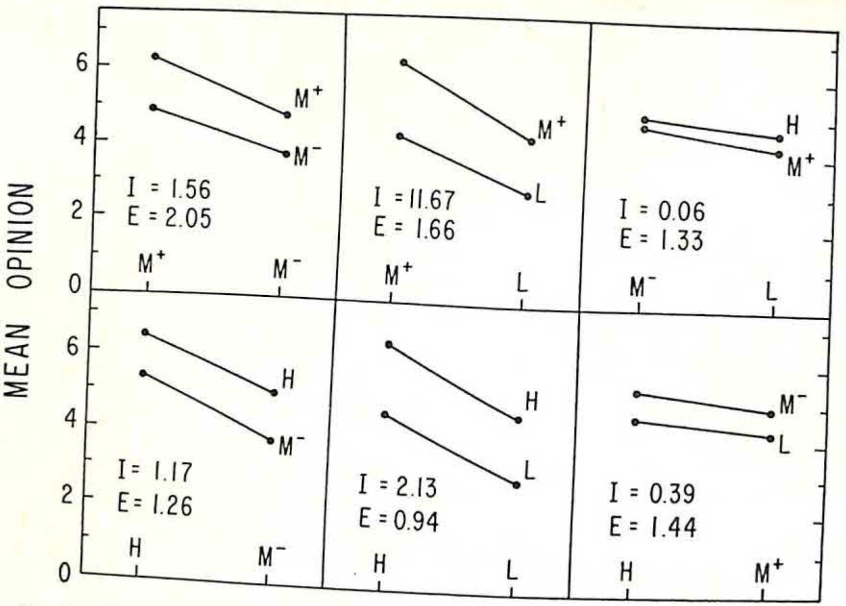


FIG. 1. Mean attitude toward United States Presidents produced by pairs of informational paragraphs. (Equal weighting model predicts parallelism for the 2 × 2 design in each panel. Mean squares for interaction and error denoted by I and E. H, M$^+$, M$^-$, and L denote paragraphs of highly favorable, moderately favorable, moderately unfavorable, and very unfavorable information.) (Reprinted from an article by Barbara K. Sawyers and Norman H. Anderson published in the May 1971 *Journal of Personality and Social Psychology*. Copyrighted by the American Psychological Association, Inc., 1971.)

scribed by two paragraphs based on historical sources (Sawyers & Anderson, 1971). Very positive, mildly positive, mildly negative, and very negative paragraphs are denoted by H, M$^+$, M$^-$, and L. Each panel shows a 2 × 2 design which should obey the parallelism prediction. Parallelism seems to hold fairly well except that the discrepancy was significant in the top center panel.

Adding and averaging models make the same parallelism prediction in this experiment and are not distinguishable from these data. However, subsequent work appears to rule out the adding model for President judgments.

The parallelism test is made directly on the raw response and does not require actual estimation of the scale values. Normative ratings were used to class the paragraphs as H, M$^+$, M$^-$, or L, but the only function of these categories was to separate the data curves in order to allow potential nonparallelism to appear. It should be noted that the 2 × 2 design is frequently not convenient or usable for functional scaling. However, the primary purpose of this experiment was to test the model rather than scale the stimuli.

Application 2: Variation in w

This application is similar to the first, but with manipulation of the weight parameter instead of the scale value. This application would be appropriate for studying the combined effects of two different sources. Since it can provide an interval scale of the weight parameter, it may be especially useful in the analysis of the many determinants of source effects.

A factorial design is employed, as in Application 1, but with the roles of w and s interchanged. Each factor corresponds to one communication; the levels within each factor correspond to different sources to which the communication is attributed. For a two-way design, the model becomes

$$R_{ij} = w_0 s_0 + w_{Ri} s_R + w_{Cj} s_C, \qquad [4]$$

parallel to Equation 2.

The assumptions of Equation 4 are analogous to those of Equation 2, though perhaps more easily satisfied in practice. The row communication is assumed to have the same scale value regardless of which source it is combined with, and similarly for the column communication. The weight of each communication depends on the source and will in general be different in each row and column. Finally, as in all these applications, it is assumed that the combining of two stimuli within any cell of the design does not change the parameter values.

For a strict adding model, the analysis is identical to that of Application 1. If the data satisfy the model, therefore, they provide equal-interval scales of the weight parameters. This application may thus be useful in the analysis of source effects, though it should be recognized that it assumes a strict adding model.

Application 3: Variation in both w *and* s

Application 1 was concerned with how two communications are combined, and Application 2 with how two sources combine. Application 3 is concerned with how one source and one communication combine. It may be particularly useful when only a single source-communication can be given.

Mathematically, Application 3 is interesting since it rests on a multiplying model instead of an adding model. As a consequence, the model analysis requires some new developments. A two-way factorial design is used again, but with the row factor representing different sources and the column factor representing different communications. The model then becomes

$$R_{ij} = w_0 s_0 + w_{Ri} s_{Cj}, \qquad [5]$$

where w_{Ri} is the weight of the source in row i, and s_{Cj} is the value of the communication in column j. In contrast to Applications 1 and 2, there is only one communication in each cell of this design. The assumptions required by Equation 5 are that the weight parameters depend only on the row and the scale values depend only on the column.

Equation 5 is a multiplying model, since the initial opinion term is just an additive constant. Functional measurement for multiplying models has been discussed for psychophysical judgment (Anderson, 1970a), and it has been used to scale subjective probability and utility in decision-making experiments (Anderson & Shanteau, 1970; Shanteau, 1970b). The same approach may be used here.

TABLE 2

Hypothetical Data for Source × Communication Design to Illustrate Functional Measurement for a Multiplying Model

Source	Communication					
	1	2	3	4	5	*M*
1	9	14	29	24	44	24
2	8	12	24	20	36	20
3	6	8	14	12	20	12
4	5	6	9	8	12	8
M	7	10	19	16	28	
Est(s_j)	0	1	4	3	7	

The hypothetical data of Table 2 will illustrate the procedure. Each entry is to be considered as the attitude resulting from one combination of source and communication. The first step is to test whether these data follow the multiplying model of Equation 5.

Test of fit. Testing the model would be straightforward if the communication scale values were known. Plotting the entries in any one row as a function of the s_{Cj} would, if Equation 5 was correct, yield a straight line with slope w_{Ri}. Although the s_{Cj} are not known, this idea provides the basis for solving the problem.

A simple graphic solution may be obtained as follows. First, two communications are assigned arbitrary values: in Table 2, the first two communications have been assigned the values 0 and 1 in the last row of the table. The observed means for these communications, 7 and 10 in the table, are plotted as a function of these assigned scale values. These two points define a straight line, and each remaining communication mean is then marked on this line at its proper elevation. The downward projection of this point on the value axis is a provisional estimate of the scale value of that communication. Thus, if 19 is marked on the line joining 7 and 10, its projection on the value axis, namely 4, is a provisional estimate of the scale value of Communication 3. In the same way, the scale values of Communications 4 and 5 are estimated as 3 and 7 in the last row of the table.

The final and critical step is to plot the data of each separate row as a function of these provisional scale values. If the model is correct, these row curves will form a family of diverging straight lines. Deviations from this pattern would infirm the model.

To supplement this graphical test, an exact statistical test is also available by appropriate use of analysis of variance. Since the model predicts nonparallel curves, the Row × Column interaction is nonzero. However, if the model is correct, all the interaction is concentrated in the Linear × Linear component, and the residual interaction should be nonsignificant (Anderson, 1970a, p. 157).

Two extensions of this application deserve mention. The first adds on a second two-way design so that each cell contains two such source-communication combinations. The analysis would then be similar to the duplex bets studied in Anderson and Shanteau (1970). Another extension adds on a third factor representing a second source variable so that each communication is attributed to two sources. This would be useful in the theoretical analysis of source combinations.

Functional scales. Verification of the model simultaneously validates the provisional scale values used in testing it. Indeed, Equation 5 implies directly that the column means of the Source × Communication design are a linear function of the communication values, s_{Cj}. Similarly, the row means are a linear function of the source-communication weight, w_{Ri}. The marginal means of the factorial design thus provide equal-interval scales of the stimulus parameters. The multiplying model is just like the adding models in this respect.

Application 4: Variation in both w *and* s

The last application is a generalization of Application 1 in which a different source-communication is allowed in each row and in each column. It may thus be applicable to situations in which the communication interacts with the source, as in studies of source expertise.

A two-way factorial design is used again. The strict adding model then implies that

$$R_{ij} = w_0 s_0 + w_{Ri} s_{Ri} + w_{Cj} s_{Cj}, \qquad [6]$$

in which different weights and values are allowed in each row and in each column. The only restrictive assumption is that the row and column stimuli do not interact when combined in any cell. However, source and communication in any given row or column may interact in any manner.

Even in this general case, the adding model still makes the parallelism prediction. Analysis of variance can thus be applied just as in Application 1. There is a loss in precision since the functional scales do not separate the weights and the scale values. Nevertheless, the row means still measure something useful, namely, $w_{Ri} s_{Ri}$, which is just the total effect of the given communication.

Applications 1 and 4 bear a close relation. In practice, the row weight in Equation 2 might depend on the scale value of the communication. It could not then be considered as a constant, w_R, but would need to be taken as a variable, w_{Ri}. The adding model predicts parallelism in either case. In an averaging model, however, the analysis becomes more difficult as will now be seen.

AVERAGING THEORY

There is increasing evidence for the importance and ubiquity of averaging processes in information integration. This finding has fundamental importance in integration theory. It affects every aspect of an experiment, from general design and procedure to theoretical interpretation. The critical evidence will be briefly noted, and then the four applications discussed above will be considered from the averaging standpoint.

Averaging versus Adding

Two main experimental paradigms have been used for critical comparisons between averaging and adding. First, adding mildly polarized to highly polarized stimulus information decreases response polarity (Anderson, 1965a; Anderson & Alexander, 1971; Hendrick, 1968; Lampel & Anderson, 1968; Manis, Gleason, & Dawes, 1966; Levin & Schmidt, 1970; Oden & Anderson, 1971). This result eliminates a direct adding formulation, though it might be saved by postulating a contrast effect. Thus, an adding model would predict a decrement if mildly favorable information received a negative evaluation in the context of highly favorable information. It seems doubtful that such contrast effects play an effective role with verbal stimuli (e.g., Anderson, 1971a, 1971b; Anderson & Lampel, 1965; Kaplan, 1971; Tesser, 1968; Wyer & Dermer, 1968), but the possibility cannot yet be ruled out.

The second result is that adding information of the same value typically makes the response more extreme; this is the familiar set-size effect. Qualitatively, the set-size effect accords with an adding process, though no adding formulation has yet given a quantitative account. More seriously, the set-size effect eliminates a simple averaging model in which a constant stimulus mean would produce a constant response.

The present averaging formulation accounts for the set-size effect by the use of an initial opinion. The set-size function is then a growth curve of somewhat different form for serial presentation (Anderson, 1959a) than for simultaneous presentation (Anderson, 1965a). In this form, the averaging model can give a good quantitative account of the set-size effect (Anderson, 1967).

This last point bears emphasis in two respects since it has sometimes been misinterpreted. First, the averaging formulation can account for both incremental and decremental effects of added information, so there is no need to postulate two processes. Second, the averaging formulation has given exact, quantitative accounts of both the incremental and decremental effects.

An interesting appearance of the set-size effect can be seen in an argument of M. J. Rosenberg (1968), who makes the plausible argument that a given piece of information that could produce considerable opinion change if the issue was new and unfamiliar would have little effect if the issue was old and familiar. Rosenberg interpreted this argument against averaging as well as adding. However, it accords with the averaging set-size function on which the old opinion, based on a considerable accumulation of information, would be near asymptote. More precisely, the old opinion, being based on much more evidence, would have a much greater weight parameter than the new opinion. In the averaging formulation, the weight of the given piece of information must then be much less relative to the old opinion, since the weights must sum to one.

Averaging Assumptions

Mathematically, the difference between adding and averaging models is simple. Adding models impose no constraints on the weight parameters, whereas averaging models require the weights to sum to one.

Averaging always implies stimulus interaction because of this constraint on the weight parameters. In an adding model, the effect of any one stimulus may be independent of the other stimuli. But in an averaging model, the effect of any one stimulus is inherently dependent on the whole set of information. In a certain sense, therefore, an averaging model has gestalt qualities, since the role of each part depends on the whole.

For present purposes, therefore, the assumption of no interaction will be used to mean that each stimulus has a natural or absolute weight which, together with its scale value, is constant across stimulus combinations. The effective weight of each stimulus is then its absolute weight divided by the absolute weights of all the stimuli in the given combination. In this way, each application considered above is directly convertible to an averaging model by dividing by the sum of the weights, and this procedure will be followed here. These effective or normalized weights sum to one for each stimulus combination. Mathematically, it is often convenient to use unnormalized weights at intermediate stages in the calculations or derivations.

Application 1

For this application, the weight w_{R} was assumed constant across rows, and the weight w_{C} was assumed constant across columns. Equation 2 then implies that the sum of the weights, $w_0 + w_{\mathrm{R}} + w_{\mathrm{C}}$, is constant, the same in each cell of the design. Dividing each predicted attitude by this constant is equivalent to changing the unit or range of the response scale. Since the scale unit is arbitrary, adding and averaging are not distinguishable under the given assumptions. Both predict parallelism, and both yield interval scales of the stimuli.

This equivalence of the two models disappears as soon as the weights vary across rows or columns. Not much is lost under the adding model; it still predicts parallelism as discussed under Application 4 above. But since the sum of the weights will now vary from cell to cell of the design, the averaging model predicts systematic deviations from parallelism.

Because of this complication in the averaging analysis, experimental precautions to help ensure equal weighting may be worthwhile. Of course, this may not be possible when the stimuli vary naturally in weight. In the personality-impression task, for instance, the response to a combination of moderate and extreme negative adjectives is displaced toward the extreme (Anderson, 1965a, 1968b; Anderson & Alexander, 1971). This directional result is consistent with the hypothesis that the extreme adjectives have greater natural weight and is one piece of evidence that favors the averaging hypothesis.

Application 2

The averaging model is quite different from the strict adding model in this design. Written as an averaging model, Equation 4 becomes

$$R_{ij} = (w_0 s_0 + w_{\mathrm{R}i} s_{\mathrm{R}} + w_{\mathrm{C}j} s_{\mathrm{C}}) / (w_0 + w_{\mathrm{R}i} + w_{\mathrm{C}j}). \quad [7]$$

Since the denominator varies from cell to cell of the design, the parallelism prediction will not hold. The deviations from parallelism will still show a systematic pattern, however, depending on the pattern of weight parameters. The expected form of the nonparallelism can be studied by using Equation 7 to determine how the difference between any two rows varies with column weight. The mathematical analysis is fairly straightforward but too detailed to include here.

To illustrate the matter, consider a square design for which the weights in row i and column i are equal. Suppose that the weights are arranged in increasing order of magnitude, from left to right across columns, and from top to bottom down rows. Finally, assume that w_0, the weight of the initial opinion, is zero. (The design provides a partial check on this last assumption, since it implies that the entries in every cell along the main diagonal are constant, equal to the simple mean value of the row and column communications. If w_0 is not zero, then averaging theory predicts a trend along the main diagonal from s_0 toward the mean of s_{R} and s_{C}.)

In this design, the global shape of the set of predicted curves is similar to a correlation ellipse, truncated at either end. Above the diagonal, the row curves converge to the right; below the diagonal, the row curves converge to the left. The overall trend is upward if the column communication has greater value, downward if the row communication has greater value. The magnitude of the trend depends on the difference between the values of the row and column communications. If these are equal, then the predicted response is the same in all cells.

A simple experimental realization of this illustrative design would be available with the personality-impression task. The row and column communications would correspond to two sets of personality-trait adjectives describing a single person. Weight could then be manipulated by specifying the occupations of the sources (Rosenbaum & Levin, 1968, 1969), for instance, or by otherwise varying the reliability of the information.

Application 3

Under the averaging hypothesis, Equation 5 becomes

$$R_{ij} = (w_0 s_0 + w_{\mathrm{R}i} s_{\mathrm{C}j}) / (w_0 + w_{\mathrm{R}i}). \quad [8]$$

If w_0 is zero, then R_{ij} equals the scale value of the communication regardless of the source weight. This might seem odd, but with no other basis for opinion, the communication must be taken at face value. Since this case is mathematically trivial, it will be assumed that w_0 is not zero.

The analysis under the averaging hypothesis parallels that of Application 3 above except in one respect. If Equation 8 is averaged over rows, the marginal column means are a linear function of the communication values. In practice, this means that the observed column means may be used as provisional estimates of the scale values in the bilinear test of fit. To the degree that the model is validated, the column means may be considered as interval scales of the communication values.

The one change required by the averaging constraint comes in the estimation of the weight parameters. In the simple multiplying model, the slopes of the bilinearized curves are proportional to the weights, $w_{\mathrm{R}i}$; with the averaging constraint, they are proportional to $w_{\mathrm{R}i}/(w_{\mathrm{R}i} + w_0)$.

Application 4

Only one case of Application 4, one that has considerable importance, will be considered in detail. The model for this case may be written,

$$R_{ij} = (w_0 s_0 + w_{\mathrm{R}} s_{\mathrm{R}i} + w_{\mathrm{C}j} s_{\mathrm{C}j})/(w_0 + w_{\mathrm{R}} + w_{\mathrm{C}j}). \quad [9]$$

This is the averaging form of Equation 6 under the added restriction that the row weight is constant.

Figure 2 illustrates the difference between the adding and averaging models. These plots are for a 3 × 4 design for which the communication scale values are those listed in Table 1B. The left panel is calculated on the assumption that the row and column weights both equal one. These three curves are parallel. The right panel is calculated on the assumption that the row weight is one, but that the column weight increases with scale value as specified in the figure legend. These curves are not parallel.

Despite this nonparallelism, the right panel of Figure 2 shows two regularities. The first is qualitative: the curves converge

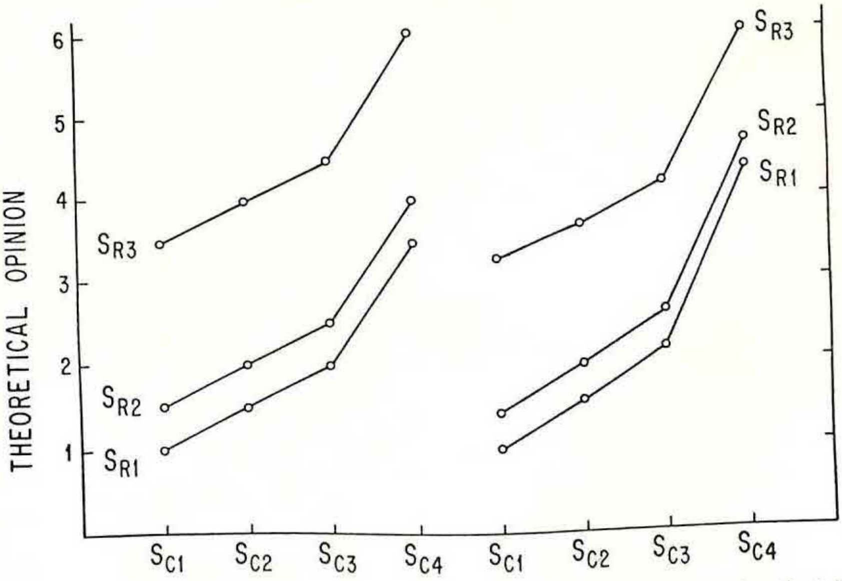


FIG. 2. Illustration of parallelism prediction with equal weighting in left graph, nonparallelism with unequal weighting in right graph. (Each point is the predicted opinion resulting from two communications whose scale values are given in Table 1B. In the right graph, the weight of the row communication is 1, while the weight of the column communication increases with scale value according to the formula, $1 + .2s_{\mathrm{C}j}$.)

toward the right, that is, as the scale value increases. This illustrates a general convergence prediction of the differential weighting model: Whenever the weights of the informational stimuli are directly related to their magnitudes, this same convergence pattern will be obtained. If the design were extended to include stimuli with negative values, this pattern would reappear as a convergence from the middle toward the left. Bilateral asymmetry would reflect differential weighting of positive and negative stimuli.

The second regularity in the right panel of Figure 2 is quantitative; the difference between any two row curves is a constant multiple of the difference between any other two row curves. This can be seen more readily by reference to Equation 9. The difference between the first two rows is

$$R_{2j} - R_{1j} = w_{\mathrm{R}}(s_{\mathrm{R}2} - s_{\mathrm{R}1})/(w_0 + w_{\mathrm{R}} + w_{\mathrm{C}j}). \quad [10]$$

If the difference between the first and third rows is expressed similarly, the ratio of these two differences is

$$(R_{2j} - R_{1j})/(R_{3j} - R_{1j}) = (s_{\mathrm{R}2} - s_{\mathrm{R}1})/(s_{\mathrm{R}3} - s_{\mathrm{R}1}), \quad [11a]$$

which is constant, independent of columns. Equation 11a can thus be rewritten,

$$(R_{2j} - R_{1j}) = a(R_{3j} - R_{1j}), \quad j = 1,2,\cdots, \quad [11b]$$

where a is a constant. This result holds in general so that any two sets of row differ-ences should plot as straight-line functions of each other. This property could be used as a rough graphical test of fit.

Another graphical test, somewhat sim-ler, can be obtained by plotting the en-ries in each column as a function of the ow means. The model implies that each olumn should plot as a straight line except r sampling error.

Two functional scales are also available. If Equation 9 is averaged over columns, it becomes a linear function of the row scale value, $s_{\mathrm{R}i}$. As in Application 1, the row marginal means provide an equal-interval scale of the values of the row communications.

Furthermore, Equation 10 indicates that the row difference is an inverse measure of the column weight, $w_{\mathrm{C}j}$. Thus, the reciprocals of the row differences can provide an interval scale of the column weight parameter as shown in Equation 12:

$$w_{\mathrm{C}j} = -(w_0 + w_{\mathrm{R}}) + w_{\mathrm{R}}(s_{\mathrm{R}1} - s_{\mathrm{R}2})/(R_{1j} - R_{2j}). \quad [12]$$

The optimal statistical procedure for estimating these weights is not known, however.

Differential Weighted Averaging

In the application just considered, the weight parameter was allowed to vary arbi-

---

An interesting extension of bilinearity analysis nderson, 1970a, p. 157) provides an exact statisti-l test. The computing procedure given there may followed exactly with these modifications. First reduced data matrix is obtained by subtracting e column mean from all entries in each column. is has the effect of reducing Equation 9 to the form of a simple multiplying model. All further calculations are done in terms of this reduced matrix. The comparison coefficients for rows are just the row means, since the grand mean is zero. To get the comparison coefficients for columns requires a detour, since all column means are zero in the reduced matrix. For this purpose only, the sign of all cell entries is ignored; the column comparison coefficients are then obtained as in the cited reference as column deviations from the grand mean in this temporary matrix of absolute values. These comparison coefficients allow calculation of the Linear × Linear component of the interaction. The residual interaction should then be nonsignificant. Its *df* equals $(r - 2)(c - 1) - 1$, where r and c are the number of rows and columns; the initial reduction of the data matrix loses one degree of freedom per column through estimating an additive constant in each column.

The error term requires a brief comment. The within-cells error would be used for both the Linear × Linear component and the Residual interaction for the case of independent groups in each cell and, typically, for single subject analysis as well. Repeated-measurements designs, with each subject in all conditions, are more complicated. If the analysis is to be made over subjects, the above procedure needs to be applied to each subject. The overall Row × Column interaction is split into Linear × Linear and Residual components using the means of the individual comparison coefficients. In parallel, the Subject × Row × Column interaction is split into Subjects × Linear × Linear, and Subjects × Residual. Each component of the overall interaction is then tested against its interaction with subjects.

trarily as a function of the scale value. There is increasing evidence that such differential weighting is more the rule than the exception (Lampel & Anderson, 1968; Oden & Anderson, 1971), and differential weighting seems especially likely in attitude change experiments. It is fortunate, therefore, that some direct analyses and estimation methods are possible.

The importance of this result can be illustrated by the date ratings of Lampel and Anderson (1968) which illustrate that a cognitive interaction of some complexity can be handled simply. Boys described by a photo and two personality-trait adjectives were rated as coffee dates by college girls. The data provided a critical test of averaging–adding in favor of averaging. In addition, the Adjective × Photo interaction was very large, with the adjectives having a greater effect with a positive than with a negative photo. This was interpreted as a discounting effect: girls do not want to be seen dating unattractive boys no matter how nice they are; as their physical appearance improves, their personality characteristics become more important. In theoretical terms, then, the weight of the photo is inversely related to its value.

In this experiment, the adjectives correspond to the rows, the photos to the columns of Application 4. Figure 1 of the cited article plots two difference curves analogous to Equation 10. The upper curve is almost exactly a constant multiple of the lower curve, in agreement with Equations 11a and 11b.

Conjunctive and disjunctive models (e.g., Coombs, 1964; Torgerson, 1958) can also be represented as differential weighted averaging models. The standard conjunctive model sets a criterion on each of several dimensions. Only if the criterion is exceeded on all dimensions simultaneously is a positive decision made. The present approach would replace the criterion by a weighting function that had high weights below some threshold region, diminishing to lower weights at higher levels of the scale value. Because the averaging model requires the weights to sum to one, any stimulus variable that fell below the criterion could dominate the response and prevent a positive decision. The averaging model thus provides a more general and flexible analysis. In particular, it replaces the somewhat artificial point criterion by a band or region.

It should also be noted that scaling is an integral part of psychological theory in the present approach. The standard treatments of conjunctive and disjunctive models, in contrast, assume that stimulus scale values are already at hand. But if weight and scale value are inversely related, as in the date-rating experiment, this critically affects the scaling analysis.

It has been suggested that a strict adding model should have been applied to the main data of the date-rating experiment. The observed interactions then would be interpreted as an artifact of a merely ordinal response scale, to be eliminated by appropriate monotone transformation (Anderson, 1962b; Bogartz & Wackwitz, 1970; Kruskal, 1965). This same argument holds more generally for the averaging analysis of Application 4. A convergence interaction could reflect a nonlinear output function (Bogartz & Wackwitz, 1970, 1971) rather than a real averaging effect. Because of their simplicity, strict adding models have been favorites in abstract measurement theory, and they have many mathematically convenient properties. However, the empirical evidence summarized under Averaging versus Adding above raises considerable doubt about the validity of strict adding models.

RELATIONS TO OTHER THEORIES

This section compares integration theory with several other theories of attitude change, mainly those that have attempted some degree of quantitative analysis. The main purpose of these comparisons is to show how integration theory handles various traditional problems in attitude theory.

Balance Theory

As an organizing concept, a principle of information integration has certain advantages over the principles of balance and congruity (Anderson, 1968a). Critical comparisons by Lindner (1970, 1971) have supported integration theory. This section will illustrate the theoretical comparison in more detail. Since the most extensive attempt to relate balance theory to attitude change is that of Feather (1964, 1967), only his development will be considered here. A recent review of balance theory is given by Zajonc (1968). Mention should also be made of the work of Wiest (1965), who has attempted to quantify the strengths of the relations, and of Newcomb (1968), who has attempted to give a more realistic definition of balanced states.

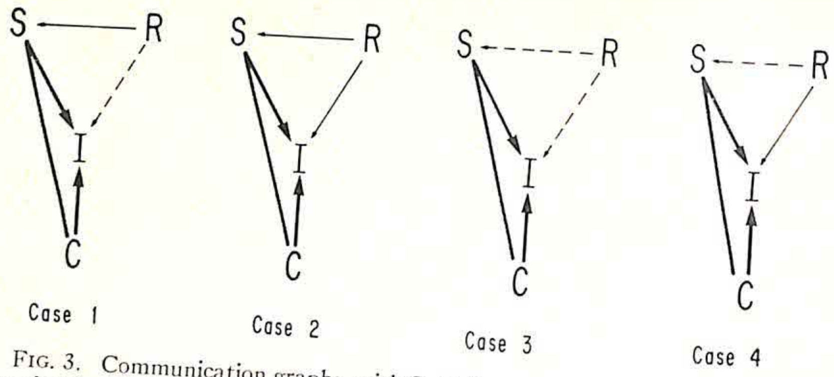


FIG. 3. Communication graphs, with S, C, I, and R denoting Source, Communication, Issue, and Receiver. (Solid lines denote a positive association or attitude; broken lines denote a negative attitude.)

Feather bases his analysis on the graph diagrams studied by Cartwright and Harary (1956) and illustrated in Figure 3. The solid arrow connecting C to I means that the communication (C) is favorable toward the issue (I) in each of the four cases. Similarly, the solid arrow connecting S to I means that the receiver (R) considers the source (S) to be favorable toward the issue prior to receiving the communication. The solid line connecting S to C simply means that the communication comes from the source. The initial attitudes of the receiver toward the issue and the source are denoted by s_0 and I_0, respectively. These features are the same in all four diagrams.

Attitude toward issue. For each diagram in Figure 3, the receiver gets two pieces of information about the issue: the initial stand of the source and the favorable communication from the source. For simplicity, it will be assumed that the prior stand of the source has already been integrated and become part of s_0. The integration equation for the attitude after receipt of the favorable communication is then

$$s = w_0 s_0 + w_1 s_C, \qquad [13]$$

a weighted sum or average of the initial attitude, s_0, and the scale value of the communication, s_C. The effective weight, w_1, of the communication will depend on properties of the source and communication as discussed elsewhere.

The four cases in Figure 3 differ according as initial attitudes of the receiver toward the source and issue are positive (solid arrow) or negative (broken arrow). These will now be discussed in turn.

Case 1: Here the receiver's initial attitude is positive toward the source, negative toward the issue. Figure 3 then represents an unbalanced structure, and balance theory postulates a tendency to restore balance. In integration theory this tendency is not a postulate but a deduction. In Equation 13, s_0 is negative, because the initial attitude toward the issue is negative, but both w_1 and s_C are positive. Accordingly, Equation 13 implies directly that the postcommunication attitude will be less negative, perhaps even positive.

This is no more than common sense, of course, but integration theory differs from balance theory in two important ways. First, the final state need not be balanced; the postcommunication attitude may be predictably negative. Second, integration theory allows a quantitative analysis of the magnitude of attitude change and its dependence on various stimulus parameters. It thus takes into account the strengths of the relations which balance theory is unable to handle satisfactorily.

Case 2: Here the receiver's initial attitudes toward source and issue are both positive. Since there is no imbalance in the diagram, balance theory has no basis on which to predict a change in attitude.

Integration theory does predict a change, following Equation 13. Under the averaging model, moreover, the changed attitude will be more positive or less positive, according as s_C, the communication value, is more or less positive than s_0, the initial attitude.

Case 3: Here the receiver is initially negative toward source and issue both as indicated by the broken arrows in Figure 3. Since the SRI triad is balanced, balance theory again lacks a basis on which to predict a change in attitude.

Under integration theory, the theoretical analysis depends on whether w_1 is positive or negative. That w_1 may be positive even with a negative source is indicated by the positive attitude change found with low credible sources (see Hovland et al., 1953; McGuire, 1969). If w_1 is positive, then the receiver's attitude toward the issue will become less negative. Indeed, it may even become positive, thus converting a balanced state to an imbalanced state.

Case 4: Here the receiver has a negative attitude toward the source but is positive toward the issue. This unbalanced state could become balanced if either the receiver adopted a positive attitude toward the source, or else a negative attitude toward the issue. Balance theory does not predict which will occur.

For integration theory, the theoretical analysis again depends on whether w_1 is positive or negative. If w_1 is negative, the receiver's attitude toward the issue will become less positive, and hence more balanced. However, if w_1 is positive, then the receiver's attitude toward the issue becomes even more positive, increasing the imbalance. As in Case 3, this allows a critical qualitative test between balance theory and integration theory.

Attitude toward source. Integration theory analyzes attitude toward the source in much the same way as just illustrated for attitude toward the issue. In the diagrams of Figure 3, there are three relevant informational stimuli, including the receiver's prior attitude toward the source. Thus, the communication attributed to the source constitutes a piece of information about the source; as such it has a certain weight and scale value. The same holds for the initial stand attributed to the source. Both these pieces of information are then to be integrated into the receiver's initial opinion about the source, and the integration process is assumed to be essentially the same as for the issue.

There are, however, important differences in detail, especially in the valuation process, that reflect the asymmetry of the source-issue relation. That the source favors the issue does not mean that the issue favors the source, for instance. In addition, the valuation process will depend on the dimension of judgment. In particular, the parameters of the informational stimuli will be different if the receiver judges the position of the source on the issue than if he makes an evaluative judgment of the source.

Evaluation of the communication. Judgments of the communication, such as its position on the issue, or its fairness, will of course depend in detail on the content and structure of the communication. They will also, as is well known, depend on the attitude of the receiver and his valuation of the source, and this dependence will be discussed briefly.

If the receiver is asked to judge the position of the communication on the issue, he has two main pieces of information to rely on. One is his opinion, explicit or inferred, about the source's stand on the issue independent of the communication. The other is the content of the communication per se. Under the simplest theoretical analysis, the judgment of the communication will then be a weighted average of these two pieces of information. Loosely speaking, the judgment might be considered a compromise between what the source actually said and what the source was expected to say; certainly this is a reasonable strategy from the subject's view.

From this standpoint, the theoretical analysis of the communication parallels that just given for judgments of the issue. For instance, if the initial attitude toward the source is negative, the communication will ordinarily be seen as less positive when attributed to the source than when considered alone. Here as before, a tendency toward balance arises as a deduction from integration theory.

This simple averaging analysis does not allow for the seeming fact that the judged position of a communication may also depend on the receiver's attitude toward the issue. This is a complicated problem (see Kiesler, Collins, & Miller, 1969, pp. 264 ff.) that will be briefly considered under Assimilation-Contrast Theory, below.

Judgments of the fairness or plausibility of the communication have also been considered (see, e.g., Brigham & Cook, 1970; McGuire, 1969). The general finding is that communications more discrepant from one's own position tend to be judged as less fair. This commonsense result merits a commonsense explanation: if the receiver is at all attached to his own position and considers it correct, he will naturally judge a discrepant communication as incorrect, hence misleading, and hence unfair. At the same time, many experimental studies of opinion change do not involve firm opinions, and in such cases fairness would not necessarily be affected. Thus, it seems more promising to look for a social learning base for such evaluations rather than a principle of balance.

Summary comments. The main conclusion from the preceding analysis is that a tendency toward balance may be derived from a principle of information integration.

Further, the integration principle is more general since it also accounts for opinion change in balanced situations. Only certain illustrative situations have been considered, of course, but much the same analysis would seem to apply to the numerous other graph structures considered by Feather (1967).

Integration theory has additional advantages. Feather (1964, 1967) has noted that major unresolved problems in balance theory include allowing for the strengths of relations, for the importance of the issue, as well as for individual differences. These problems are handled in a natural way, in terms of the w and s parameters of integration theory. Moreover, functional measurement technique is specifically oriented toward measurement at the level of the individual.

Heider (1958) has stated that "Formulation in terms of equifinality [balanced end state] is more parsimonious than formulation in terms of single conditions and effects [p. 207]." Integration theory takes the opposite road to parsimony, stressing the primary importance of detailed analysis, both theoretical and experimental, of stimulus conditions and their effects.

Congruity Theory

Congruity theory (Osgood & Tannenbaum, 1955; Tannenbaum, 1967, 1968) postulates an averaging model of the form of Equation 1 (S. Rosenberg, 1968). It is, however, a very specialized averaging model, since it requires the weight parameters to be specific functions of the scale values or polarities of the stimuli. For two stimuli of values s_1 and s_2, congruity theory postulates that $w_1 = |s_1|/(|s_1| + |s_2|)$, $w_2 = |s_2|/(|s_1| + |s_2|)$. This formulation makes the weight increase with the scale value, which would be reasonable in many situations. It also requires a neutral stimulus to have zero weight, which is clearly wrong.

In the present formulation, the weight parameter can depend on factors other than stimulus value. For instance, one can experimentally manipulate the credibility or reliability of the information without thereby affecting its scale value. This indicates that congruity theory has an inadequate conceptual base. Moreover, congruity theory has not been able to handle the set-size effect, in which the addition of information of equal value increases the extremity of the response (e.g., Anderson, 1959a, 1965a, 1967). An extensive discussion of this and related problems is given by Smith (1970).

Tannenbaum (1967) has recently discussed two experiments that he interprets to support congruity theory and to infirm an information-processing approach. However, both experiments can be treated as 2 × 2 designs for which integration theory predicts zero interaction, in agreement with the data. An earlier experiment by Tannenbaum and Norris (1965) also showed a nonsignificant interaction, though there the theoretical analysis is less clear.

In one experiment (Tannenbaum, 1967, Table VI), each condition contained one message from a positive source, one message from a negative source. The four conditions differed in that they contained two positive messages, two negative messages, or one of each polarity. The interaction term would apparently not have approached significance, in agreement with integration theory. The data imply a negative weight for the negative source, though this may reflect the covariance adjustment.

In the other experiment, the four experimental conditions received either the same or different messages, from either the same or different sources. In an informational analysis, redundant information requires a redundancy parameter, in this case for same-versus-different message as well as for same-versus-different source. These redundancy parameters need not be evaluated, but can be incorporated directly in the formal model. The adding model then predicts zero interaction, and the averaging model makes approximately the same prediction. Tannenbaum's data (1967, Table V) show a very small interaction term that would evidently not have been significant.

Neither experiment, therefore, poses any difficulty for information integration theory. Indeed, integration theory goes beyond congruity theory to make quantitative predictions in both cited experiments.

Integration theory implies that congruity, like balance, will generally increase. A congruity principle, or at least a congruity tendency, is thus a deduction from a more general principle of information integration. The detailed supporting analysis for this claim has already been given in the discussion of balance theory. Here it may be noted that integration theory avoids a fundamental criticism raised by Abelson (1963). If a positive source makes a positive statement about some issue, the evaluation of source and issue will approach each other. The cause, according to congruity theory, is a pressure toward congruity which exists as long as the source and issue evaluations are unequal. This makes it awkward to explain why complete congruity is seldom if ever obtained. For integration theory, there is no difficulty. The change is effected by the information in the communication; once this is absorbed, no further change is implied and there is no reason to expect complete congruity between source and issue.

Assimilation-Contrast Theory

The present theory of opinions and attitudes is in large part a theory of social judgment. Although this section is meant to compare integration theory with other quantitative formulations, comparison with assimilation-contrast theory (Sherif & Hovland, 1961; Sherif, Sherif, & Nebergall, 1965) is appropriate since it also emphasizes judgmental processes. Summaries and evaluations of their theory have been given by Insko (1967) and Kiesler et al. (1969). Here the two approaches will be compared in the context of an interesting recent experiment by Rhine and Severance (1970).

Rhine and Severance studied three variables: source credibility, ego involvement, and discrepancy between the person's initial attitude and the position advocated by the communication. For a single communication with weight w_1 and value s_1, the postcommunication attitude can be written as

$$R = w_0 s_0 + w_1 s_1, \qquad [14]$$

where s_0 is the initial opinion, with weight w_0. In an averaging model, the weights are required to sum to one. This effect can be achieved by letting w_1, and $w_0 = 1 - w_1$ be the relative weights. Then Equation 14 can be rewritten as

$$R = s_0 + w_1(s_1 - s_0). \qquad [15]$$

From Equation 15, the change in attitude is seen to be just $w_1(s_1 - s_0)$. Accordingly the theoretical analysis devolves on the determinants of w_1 and of $(s_1 - s_0)$.

Source credibility will affect w_1. A more credible source would correspond directly to a larger value of w_1 and would produce more attitude change.

Ego involvement was defined by choice of issue, either the proposed tuition increase in the University of California, Riverside, where the experiment was run, or the desired size of the city park in Allentown, Pennsylvania. For this particular manipulation, ego involvement corresponds to w_0, the strength of the initial opinion. Thus, w_1 would be lower for the higher level of ego involvement and less change would occur.

The discrepancy variable, $(s_1 - s_0)$, was manipulated by using three different dollar or acre values in the communications given to different groups. With no further qualification, Equation 15 predicts that amount of opinion change will be a linear function of the discrepancy, $(s_1 - s_0)$. This was found to hold for the parks issue though not on the tuition issue.

Integration theory includes a process of inconsistency discounting (Anderson & Jacobson, 1965) in which inconsistent information is given decreased weight. For example, a communication attributed to a highly credible source would tend to be discounted if it was inconsistent with the person's previous opinion about the source's position. In the same way, a communication will tend to be discounted when it is inconsistent with fact or with the person's own previous opinion on the given issue.

Accordingly, the predicted effect of the discrepancy variable needs to be considered in the light of the experimental task. For the parks issue, the subjects were told that the average park size was 20 acres for a city with a little less than half the 106,400 population of Allentown. The highest advocated figure was 240 acres which would not be expected to evoke any inconsistency reaction. On that basis, the obtained linear relation between advocated change and obtained change is in accord with expectation (Anderson, 1959a; Anderson & Hovland, 1957). Theoretically, linearity is a direct consequence of the averaging hypothesis.

The highest advocated increase in tuition was $600 which was very high, especially in view of the long tradition of free tuition that prevailed prior to the present state administration. For the tuition issue, w_1 would decrease as $(s_1 - s_0)$ increased. The predicted change, being the product of these two factors, would accordingly increase at first and later decrease as s_1 became more discrepant from s_0.

The inverted-U relation obtained for the tuition issue by Rhine and Severance, as well as similar results obtained by others (e.g., Aronson, Turner, & Carlsmith, 1963; Bochner & Insko, 1966), is thus consistent with integration theory. It should be noted, however, that a firm prediction of the later decrease depends on having sufficiently large values of the discrepancy variable. Rhine (personal communication, 1970) has suggested that this was the case on the ground that the discrepancies extended well into the latitude of rejection.

In their Table 5, Rhine and Severance list five predictions about attitude change and conclude that assimilation-contrast theory accounts for three of these while dissonance theory accounts for two. Integration theory, on the analysis just given, agrees with the data in all five cases. This analysis is post hoc, of course, but it is not ad hoc; no special assumptions were made in applying the integration model. Rhine and Severance also provide data on source and message evaluation that may be analyzed in the manner indicated in the section on balance theory.

Inconsistency discounting is central i the present analysis of the discrepancy variable. Almost all other theories of attitud change postulate some process that has th same general effect (see Insko, 1967, Ch 3). None of these is too satisfactory, since their application depends more or less on common sense. Integration theory has one advantage, since it can make definite quantitative predictions about the effects of combining two communications.

The heavy dependence of the Sherif-Hovland approach on concepts of assimilation and contrast is of theoretical concern for several reasons (see Insko, 1967; Kiesler et al., 1969; Upshaw, 1969). Assimilation-contrast is considered again later, and only two points will be made here. Both bear on a problem mentioned in the discussion of balance theory; namely, the effect of a person's own opinion on his judgment of the position of a communication.

Sherif and Hovland claim that people with strong opinions tend to see communications that portray markedly different opinions as even more discrepant than they "really" are, a contrast effect. Such claims are beset by interpretational difficulties. Defining the "real" position of the communication in terms of responses of other people is a dubious procedure. Within functional measurement theory, the scale value of the communication is unique to each person; that one person's evaluation differs from that of another does not mean that either has distorted the position of the communication. This view is consistent with Upshaw's (1965, 1969) explanation in terms of a variable zero point in the judgment scale. In general, comparisons between different populations can be most difficult (Anderson, 1963), and this is no less true when the populations are single people.

It is also possible that such displacement effects represent composite judgmental processes. For example, the overt judgment of the communication may be a weighted average of its position on the issue, and of the felt position, or perhaps of the social evaluation of the person to whom the statement is attributed. For many political and social issues, such a judgmental process would appear to produce "contrast" for positions far from that held by the judge, and "assimilation" for positions near that held by the judge. On this basis, apparent contrast in evaluation of discrepant communications would be expected from Dawes' (1971) finding that persons with opposite points of view are thought to be more extreme than they are in fact. A related model for an assimilation effect has been applied to component judgments in impression formation with some success (Anderson, 1966a; Kaplan, 1971).

Similarity-Attraction Model

An extended series of experiments by Byrne and his associates (see Byrne, 1969) has supported, under certain experimental conditions, the hypothesis that interpersonal attraction is a linear function of proportion of similar attitudes. In the typical experiment, the subject receives a form purporting to describe the attitudes of a stranger on political affiliation, drinking, and similar issues. These forms are constructed to have specified numbers of attitudes that are similar and dissimilar to the attitudes of the subject. The subject's response is a rating of how much he would like the stranger.

This interpersonal attraction task is quite similar to the personality-impression task studied by the writer (Anderson, 1967; Byrne, Lamberth, Palmer, & London, 1969). It is appropriate, therefore, to ask how integration theory might apply to the interpersonal-attraction task. Of the numerous interesting results, only one or two can be considered here.

From the view of information integration theory, each item on the form constitutes a piece of information about the stranger, much the same as a trait adjective. As such, it has a weight and a scale value along the dimension of judgment. Let w_p and s_p be the weight and value of an item marked similar, and let w_n and s_n be the weight and value of an item marked dissimilar. Suppose there are N items altogether of which k are marked similar. Under the adding model, the theoretical expression for the attraction response is then

$$R = kw_p s_p + (N - k)w_n s_n. \qquad [16]$$

If the number of items marked similar is increased by one, k would be replaced by $k + 1$ in this equation. The change in response is then the difference,

$$\Delta R = w_p s_p - w_n s_n. \qquad [17]$$

Since ΔR does not depend on k, it follows that R is a linear function of k, and hence also of k/N, the proportion of similar items. Under the given assumptions, therefore, Byrne's linearity result is consistent with prediction from integration theory. More direct empirical support has been obtained by Kaplan and Olczak (1970) who employed a factorial integration design as in Application 1.

The adding model of Equation 16 rests on the assumption that the similar items have equal value and weight, and so also for the dissimilar items. If these assumptions do not hold, the linearity prediction does not follow. However, it would still be possible to use Application 4 of the adding model to get a parallelism prediction.

If the averaging model holds, the right side of Equation 16 must be divided by the sum of the weights, $kw_p + (N - k)w_n$. The linearity prediction then holds only if $w_p = w_n$. Deviations from linearity would then be expected to the degree that similar and dissimilar items are differentially weighted. This possibility deserves further consideration since the personality-trait studies argue for an averaging model.

Two important differences between integration theory and Byrne's formulation can be illustrated in the Byrne and Rhamey (1965) experiment. In addition to the information about the stranger's own attitude, the subjects also received information about the stranger's evaluation of the subject's own intelligence, adjustment, etc. To handle two sources of information, the attraction equation was modified so that

> attraction toward X is a positive linear function of the sum of weighted positive reinforcements (Number × Magnitude) received from X divided by the total number of weighted positive and negative reinforcements received from X [Byrne & Rhamey, 1965, p. 887].

This differs from the integration model in several respects. For instance, the scale values, which correspond to their "reinforcements," do not appear in the denominator of the integration equation.

The Byrne and Rhamey experiment has special interest, since it employed a two-way design in which the interaction was significant, contrary to the simple integration model. Inspection of their data suggests that the interaction results from a discounting of the attitude-similarity information in the context of the favorable evaluation of the subject. The value of their control condition in supporting this interpretation deserves particular mention as a useful methodological device.

It should be noted that Byrne and Rhamey also predict no interaction (see Table 5, Byrne & Rhamey, 1965). That they got a close fit to the linear function despite the interaction illustrates the limitations of regression analysis discussed in the next section.

The second difference between integration theory and Byrne's formulation is more fundamental. As illustrated in the above quotation, Byrne considers the attitude-similarity items to be reinforcing stimuli. Integration theory considers them to be informational stimuli, a view that is supported by the work of Himmelfarb and Senn (1969) on judgments of social class.

This distinction between informational stimuli and reinforcing stimuli would hold no less in real social situations. From the present view, giving praise or blame does not necessarily act as a reinforcing stimulus for the attraction response to the source; it may simply provide an informational stimulus which is integrated into the opinion about the source. In applications to learning (Anderson, 1969b; Anderson & Hubert, 1963; Friedman et al., 1968), integration theory thus appears as a cognitive, information-processing theory somewhat in the sense of Tolman rather than an S–R reinforcement theory in the sense of Hull or Skinner.

Summation Theories

Certain kinds of integration tasks evidently require strict adding rather than averaging models. For example, the value of a commodity bundle might be taken as the sum of the values of its components (e.g., Bock & Jones, 1968; Gulliksen, 1956), possibly with a law of diminishing returns (Shanteau, 1970b). The addition of one commodity to the bundle would thus increase the total value. For attitude change also, an adding process is intuitively attractive, on the analogous argument that the addition of favorable information should increase the favorableness of the response.

The most vocal attempt to support an adding model for attitude change has been made by Fishbein and his associates. Triandis and Fishbein (1963) used a pe sonality-impression task to compare summation model with congruity theor Mean correlations between observed a predicted were .65 for the summation mode .53 for congruity theory. Two later e periments (L. R. Anderson & Fishbei 1965; Fishbein & Hunter, 1964), also o personality impression formation, showe similar support for a regression model ove the congruity model.

For some reason, this disconfirmatio of congruity theory was generalized to in clude all averaging models on the one hand and all balance-consistency theories on the other. But congruity theory is a very specialized averaging model as already noted. Moreover, general balance theory carries no commitment to adding or averaging (Abelson, 1968a, p. 123; Anderson 1968a). M. J. Rosenberg (1968) also discusses this matter.

One further piece of evidence was given by Fishbein and Hunter (1964). With serial presentation of personality adjectives, they found that the addition of adjectives of essentially equal value increased the extremity of the response. This set-size effect contradicts the simplest form of an averaging model. However, essentially the same result had been obtained in previous work on attitude change (Anderson, 1959a) and was interpreted to result from the averaging in of an initial opinion. Later work has supported this interpretation (e.g., Anderson 1965a, 1967; Anderson & Alexander, 1971; Hendrick, 1968; Oden & Anderson, 1971).

It should be emphasized that a quantitative model can be used to good purpose in an essentially qualitative way as exemplified by the summation formulation used by M. J. Rosenberg and by Peak. Their means–ends formulation looks at the consequences or properties of an issue or object. Each consequence is considered to have a certain value for the person (analogous to the present *s* parameter), and also a "perceived instrumentality" (Rosenberg, 1960) or "judged probability" of occurrence (Peak, 1955) (analogous to the present *w* parameter). Attitude toward the issue or object is then simply Σws, an adding formulation that seems to be conceptually similar to that used by Fishbein (see Insko, 1967, pp. 136, 197). Rosenberg tested this formulation by using posthypnotic suggestion to change attitudes on certain issues, and found associated changes in the corresponding w and s values. Although this work has been criticized on the ground that hypnosis is even less understood than attitude change, there is no doubt about its great interest and potential.

Regression-Correlation Formulations

Much work on information integration, especially in clinical judgment, has rested on multiple regression and correlation analysis (e.g., Goldberg, 1968; Hammond, Hursch, & Todd, 1964; Hoffman, 1960; Meehl, 1954, 1960; Triandis & Fishbein, 1963; Tucker, 1960; Wishner, 1960; Wyer, 1969). This approach has suffered from two serious shortcomings.

In the first place, the correlation-regression analyses that are reported seldom include a test of fit of the linear regression model. It might seem that high correlations between predicted and observed are relevant evidence. However, very high correlations are virtually guaranteed by the usual experimental designs, even with a seriously defective model (Anderson, 1962a). Adequate assessment must attend to the discrepancies from the model predictions.

That discrepancies from an adding model can be important despite high correlations is illustrated in Sidowski and Anderson (1967). Subjects judged the desirability of working at certain occupations in certain cities. An adding model was applied, and the correlations between predicted and observed were .986 and .987 in two experiments; yet there was a sizable interaction localized at the occupation of teacher in the least desirable city. Similarly, the model-data correlation for the date ratings discussed above (Lampel & Anderson, 1968) was .985 despite a strong Adjective × Photo interaction. Even in an experiment designed specifically to produce a large configural effect (Anderson & Jacobson, 1965, Condition 2), the correlation was .977. In all these experiments, the discrepancy was clearly visible as a nonparallelism, and was readily picked up by the analysis of variance.

Thus, the present approach leads to a different view from that of Goldberg's (1968) review of linear models in clinical judgment. Goldberg considers three interpretations of the general finding that human judges are seldom better than linear prediction models, and suggests the most attractive interpretation to be that judges really behave configurally but that the linear regression model is so powerful that it obscures the real configural processes. The results cited in the previous paragraph, as well as related work, point to a fourth interpretation: A linear-type integration model holds in many situations, but there are certain conditions that produce configural response; further, analysis of variance has ample power to detect configurality when it occurs (Anderson, 1969a; Slovic, 1969).

The second difficulty with regression-correlation is more fundamental. Regression analysis typically proceeds under two scaling assumptions: that the stimulus values are known, and that the overt response is on an equal-interval scale. These assumptions naturally come under suspicion if the linear regression model does not fit well. Nonlinearity may be nothing more than an inappropriate stimulus scale. For if the ostensible stimulus scale is not equal interval, then a linear function of the true subjective values will appear to be nonlinear. Similarly, apparent configurality may not represent true interaction but only an inappropriate response scale (Anderson, 1961a, 1962b; Bogartz & Wackwitz, 1970). Without a theory of measurement to support the scaling assumptions, interpretations in terms of nonlinearity or configurality rest on uncertain ground. The advantage of functional measurement theory is that it gives a unified approach to these problems.

This brief discussion should not obscure the great usefulness of regression analysis in certain practical prediction problems.

As Yntema and Torgerson (1961) indicate, systematic discrepancies from a simple additive rule of combination may not be serious in certain man–machine systems. Much of the work on clinical judgment can be viewed from a similar standpoint of practical prediction. Unfortunately, there seems to be an overwhelming temptation to generalize from product to process, that is, from prediction to understanding. Good prediction does not imply good understanding, as the correlation coefficients quoted in the third paragraph of this subsection show. When the main concern is to understand the psychological processes that underlie the judgment, any systematic discrepancy may be meaningful and important.

It should also be emphasized that it is possible to apply regression analysis to the question of goodness of fit, as well as to the measurement question. Analysis of variance and multiple regression are both applications of the general linear hypothesis of mathematical statistics and have considerable similarity. However, the typical regression design corresponds to a highly confounded factorial design; therefore, discrepancies from prediction are generally difficult to localize and interpret. Moreover, nonindependence of the beta weights (Darlington, 1968) can complicate the stimulus scaling. On the other hand, regression designs can be very efficient, since they require far fewer observations than a complete factorial design. An attempt to apply multiple regression to attitude measurement has been made by Ramsay and Case (1970), though their approach assumes that the stimulus values are known and fails to supply a test of fit. Some work on extending functional measurement to regression analysis has been done by Bogartz and Wackwitz (1970, 1971).

Logical-Consistency Theory

That human reasoning does not obey formal syllogistic logic is well known. It is also well known that formal syllogisms are only a small part of logic, so perhaps human reasoning might accord with some more general logic model. This interesting possibility has been studied by McGuire in a series of articles (e.g., McGuire, 1960, 1968) on a logic model for probabilistic beliefs. The basic postulate is that subjective probabilities or beliefs must be consistent with one another in such a way that they obey the laws of mathematical probability theory, though they may otherwise have any relation to the objective probabilities.

McGuire has applied this logic model in both qualitative and quantitative ways. The qualitative approach assumes that the tendency toward logical consistency is complicated by other factors, such as wishful thinking, and is concerned with Socratic effects, cognitive structure, and related problems. The quantitative approach, which has been adopted by Wyer and Goldberg (1970) and Wyer (1970), assumes the basic postulate, that subjective probabilities obey mathematical probability theory, is exactly true. Only this quantitative approach will be considered here.

To facilitate comparison with integration theory, only the following law of probability theory will be considered:

$$P_B = P_A P_{B/A} + P_{A'} P_{B/A'}. \qquad [18]$$

Here P_A, $P_{A'}$, and P_B are the probabilities of A, not-A, and B; $P_{B/A}$ and $P_{B/A'}$ are the conditional probabilities of B, given A and not-A. The logical-consistency model does not require subjective probabilities to have any relation to objective probabilities (e.g., expectation of success need not equal its true probability). But it does require that subjective probabilities be consistent among themselves in the manner specified by Equation 18.

That one's beliefs on any given, restricted set of issues ordinarily exhibit a fair degree of consistency is a matter of common observation. But whether such a set of beliefs will obey Equation 18 of the logic model is not easy to determine. As a precaution, the investigation might be limited to belief systems that the person himself felt to be consistent. When some new fact or argument threatens a theoretical position, the belief system may be at least temporarily deranged.

Naturalistic observation. Testing the logic model in a naturalistic setting rests completely on the measurement of the subjective probabilities. If the five separate terms of Equation 18 can be measured, then it is fairly straightforward to test between P_B as observed and predicted. An absolute measurement scale is required, however, and it is unclear that the rating scales that have been used in the experimental work are even equal-interval scales. Shanteau (1970b) has found sizable differences between probability ratings of words such as *probably*, *unlikely*, etc., and their functional scale values. Correlation-regression analysis has the same shortcomings in this work that were discussed previously.

The use of ratings as measures of subjective probability seems especially problematic for the conditional events that occur in the logic model. The implication, "If abortion is infanticide, then it should be prosecuted like other murders," (if A, then B) presumably ought to get a high $P_{B/A}$ rating; but probably such ratings would be strongly affected in many people by their belief in the premise or the conclusion (Lefford, 1946; McGuire, 1968).

Equation 18, unfortunately, may be insensitive to such biases when the beliefs are extreme. If P_A, the belief in the premise, is near zero, bias in $P_{B/A}$ will have little effect. This illustrates that the logic model may yield reasonably good predictions even when it is seriously wrong. Of course, even if the ratings of the conditional events are not valid, that does not necessarily infirm the logic model. The person's internal cogitation may be perfectly consistent with Equation 18 without his being able to assign valid ratings to the events.

Two brief comments should perhaps be added on the use of such probability ratings. First, the processes that underlie such ratings may be of considerable interest and worth closer study in their own right. Second, there are many purposes for which an equal-interval scale is not needed. Random assignment allows comparison between different experimental treatments even with an ordinal scale (Anderson, 1961a, 1963). Similarly, McGuire's (1960) interesting work on the Socratic effect does not necessarily require an interval scale.

Experimental analysis. If integration theory can be applied to these probability judgments, a radical simplification of the measurement problem is possible. Accordingly, the blanket assumption is made that all five terms in Equation 18 follow this formulation. Since P_A, $P_{A'}$, and P_B presumably follow an averaging rather than an adding model, some care would be desirable to ensure equal weighting of the communications. For the two conditionals, $P_{B/A}$ and $P_{B/A'}$, a multiplying model would seem appropriate, though the overt ratings themselves might follow an averaging model.

The main question to be considered is whether the logic model and the integration formulation are mutually consistent. Application 1, in which two communications are given in a factorial design, will be used for this purpose. This reflects a key experimental change from the work of McGuire (1960) and of Wyer and Goldberg (1970); the use of two communications in a factorial design provides the leverage for a critical comparison.

Two cases need to be considered. The first case rests on the assumption that the communications have no effect on the conditionals, $P_{B/A}$ or $P_{B/A'}$, but only affect belief in A, A', and B. By assumption, these beliefs follow the integration model and substitution of the theoretical expression from Equation 2 into Equation 18 yields the following predictions for P_B:

$$P_{Bij} = (w_0 P_{A_0} + w_R s_{Ri} + w_C s_{Cj}) P_{B/A} + (w_0' P_{A'_0} + w_R' s_{Ri}' + w_C' s_{Cj}') P_{B/A'}. \quad [19]$$

This expression is additive in the manipulated scale values. The joint application of the logic model and integration theory thus implies zero interaction in these predicted values of P_B. Since this agrees with the assumption that the judgments of P_B obey the integration model, the logic model is consistent with integration theory in this case.

This test, it should be emphasized, can be made directly on the raw ratings of P_B. The problems of measuring the four terms on the right of Equation 18 have been completely bypassed. Furthermore, because of the response scaling feature of functional measurement, only an ordinal response scale is strictly necessary for judgments of P_B. This test does depend on getting communications that do not affect the conditional probabilities, but that should be straightforward, at least if the logic model is correct.

As the second case, suppose that one or both communications also affect belief in $P_{B/A}$ and/or $P_{B/A'}$. This could be arranged experimentally by an appropriate mixture of arguments. Equation 19 must then be changed by substituting in these theoretical expressions as well. This done, the predicted values of P_{Bij} no longer obey the parallelism prediction. This contradicts the original assumption that the averaging model applies to P_B. In this second case, therefore, integration theory and the logic model are inconsistent: if P_B obeys the averaging model, the logic model cannot hold.

Both cases are needed for a proper critical test. Although the first case does not distinguish between the models, it is necessary to show that at least one of the models may be correct. The personality adjective task has various advantages as an experimental situation. The conditional probabilities would then depend on the implicational similarity of the traits.

One essential difference between the two approaches is in the conceptualization of the attitude change process for P_B. In the integration model, the informational stimuli are considered to bear directly on P_B. In the logic model, these changes occur indirectly, mediated by changes in P_A, $P_{B/A}$, etc. In the logic model, the exact analysis of Equation 19 depends on whether $P_{B/A}$ and $P_{B/A'}$ follow an averaging, adding, or compound averaging–adding model, but some simple predictions about matrix rank are available.

Bayesian statistics. Modern developments in mathematical statistics have emphasized personal probability and Bayesian statistics (Edwards, Lindman, & Savage, 1963; Savage, 1957). The Bayesian approach has close parallels with logical-consistency theory. Part of its interest lies in its emphasis on the revision of statistical probability in the light of new evidence, analogous to opinion change. Although the Bayesian development is a strictly statistical theory, considerable experimental work has been done using it as a normative model of how men ought to process probabilistic information (see Edwards, 1968; Peterson & Beach, 1967; Slovic & Lichtenstein, 1970).

It is the fate of normative theories that men do not behave as they normatively ought. Aside from the ubiquitous order effects, the response is usually much more "conservative" than is statistically correct. Several explanations of "conservatism" have been given (Edwards, 1968), including misperception, misaggregation, and response bias.

In terms of integration theory, misperception and misaggregation would both correspond to weight parameters different from the statistically optimal values. Since integration theory is descriptive rather than normative, it is directly concerned with the determinants of the weights (Shanteau, 1970a). Nonoptimal weighting in the Bayesian binomial inference task would be no more surprising than in number averaging (Anderson, 1964d, 1968d), but deviations from a statistical standard, however important they might be in practical decision problems, seem unlikely to say much about psychological processes. In the present approach, accordingly, the emphasis is on getting a theoretical framework that is psychologically congruent to the behavior.

The response bias interpretation has recently been supported by DuCharme (1970). Of itself, response bias would not necessarily be serious, since the methods of functional measurement could be used to rectify a distorted response scale. However, the work of Shanteau (1970a) suggests a more fundamental response problem. The standard bookbag-and-poker-chip task used in the Bayesian research asks for inference judgments of the probability that there are more white than red chips in a population sampled one chip at a time. Shanteau compared such inference judgments with estimation judgments about the proportion of white chips in the population. No difference was found between these two conditions, and Shanteau's careful experimental work indicated that both instruction conditions produced estimation judgments. Shanteau's results may thus invalidate the Bayesian interpretation of the binomial inference experiments.

Proportional-Change Model

The proportional-change model for attitude change (Anderson, 1959a; Anderson & Hovland, 1957) postulates that the amount of attitude change is proportional to the advocated change:

$$s_1 = s_0 + w(s_C - s_0) = ws_C + (1 - w)s_0, \quad [20]$$

where s_0 and s_1 are the opinions before and after receipt of a communication with value s_C and weight w. An analogous postulate was used in learning theory by Hull (1943). An extensive mathematical development was given by Estes and Burke (1953) and Bush and Mosteller (1955) for discrete choice tasks, and by Anderson (1961b, 1964a, 1964b) and Rouanet and Rosenberg (1964) for continuous response tasks. In these models, the weight parameter represents the learning rate.

In its original form, the proportional-change model has not been very successful. In learning, the critical tests are the sequential dependencies (Anderson, 1959b; Atkinson & Estes, 1963). The failure of stimulus sampling theory of probability learning has resulted from its inability to account for these dependencies (see e.g., Anderson, 1960, 1964c, 1966b; Anderson & Grant, 1957, 1958; Anderson & Whalen, 1960; Friedman et al., 1968; Jones, 1971).

Attitude change experiments, in contrast to learning experiments, typically involve only a few "trials." As a consequence, the weight parameter will change over successive communications (Anderson, 1959a). A related difficulty arises in Abelson's (1964) generalization of the Anderson-Hovland model to group interaction situations which, under fairly general conditions, predicts that everyone in a given group will eventually reach the same opinion. This difficulty can be avoided if the communications are treated as informational stimuli. A given communication then loses its effectiveness over successive repetitions since it conveys no new information.

The present serial integration model includes the proportional-change model as a special case and has considerably greater flexibility. A principal advantage is its ability to allow for changes in the weight parameter over a sequence of informational stimuli. Complete serial position curves, which may include both primary and recency components, thus become available (e.g., Anderson, 1964d, 1965b; Shanteau, 1970a; Weiss & Anderson, 1969).

SOURCE AND COMMUNICATION PARAMETERS

The weight and value parameters are properties of the source-communication combination. Much work on attitude change is not concerned with source effects and there may not even be an explicit source. When paragraphs about United States Presidents are given as historical fact (Sawyers & Anderson, 1971), the concept of source is relevant, but it resides at large in the cultural-experimental context. At the other extreme, the source may be explicit and its nature an integral part of the communication.

The valuation problem is interesting in its own right, and it is also relevant to the various design applications that have been considered. The problem is complex, depending heavily on the situational details, and a brief discussion necessitates somewhat cavalier treatment. The following rough analysis should be useful in many cases and is given to illustrate some of the main points without explicit concern for its evident limitations.

Molar Analysis

To begin, it seems reasonable to expect that within limits the scale value will be independent of the source. There are important exceptions, of course, especially when the source is an integral part of the communication. In oral presentations, for example, warmth of voice and gesture could become attributes of the communication and affect its scale value. However, the scale value corresponds to the position of the communication on the issue. In a substantial part of attitude change research, this position is conveyed by the semantic content of the communication and should be independent of the source.

Determination of the weight parameter is markedly more complex as shown by numerous experimental studies summarized in McGuire's canonical chapter on attitude change (McGuire, 1969, pp. 177 ff.). Some communications, based on consistency or affective arguments, can have a completely source-free weight, but in most cases the weight will depend importantly on the source. The source may be merely implicit, defined by the cultural-experimental situation, or source weight may be explicitly varied. The weight will also depend on specific properties of the communication such as clarity and redundancy.

Both the source and the communication will thus contribute to the weight parameter. If these contributions are independent, as might be expected in some situations, it would be of interest to know how they combine. The most plausible hypothesis would seem to be a multiplying model, though a composite adding-multiplying model might prove necessary.

Of course, source and communication may interact with each other, and with the issue as well, in determining the weight parameter. A communication at variance with a source's known position on the issue might be partially discredited or discounted, that is, given a lower weight. Source expertise would also interact with the issue; an engineer would presumably be more effective on pollution control than on narcotics addiction.

Application 4 of the averaging model may be especially useful in the analysis of source effects and interactions. For example, a given communication could be attributed to different sources in different columns of the design. In each case, source and communication could interact in any way. The interval scale of the weight parameter would then provide an assessment of the source-communication-issue interactions as reflected in the weight parameter.

Molecular Analysis

Thus far, the communication has been treated as a unit with a single weight and scale value. However, it will usually have a more or less complex structure, containing various separate statements and arguments, and its molar effect will itself result from information integration. The following classification is neither novel nor complete, but it indicates how a molecular analysis might proceed.

Means–ends assertions constitute one great class of persuasive arguments (see McGuire, 1969, pp. 153 ff.). The source asserts that a certain action or belief will produce a good or avoid an evil, thereby modifying the acceptance of that action or belief. Each such argument is a subcommunication, and indeed a source-communication-issue instance. The issue is the good or evil end, and is the primary determinant of the value of the subcommunication. The weight parameter will reflect the importance of the end, and also the persuasive properties of the source and communication that link means and end. On this analysis, which is similar to that of Peak (1955) and Rosenberg (1956), each means–end argument counts as one piece of information to be integrated into the overall opinion.

The other class of persuasive arguments can be called inferences. Letters of reference, for example, are usually amplified lists of personality adjectives. To judge the intelligence or motivation of someone described as level-headed requires an uncertain inference about the parameters of the given information with respect to the specified dimension of judgment. Both weight and scale value would be determined primarily by the semantic-actuarial relations between level-headedness and intelligence or motivation. Source characteristics would also affect the valuation process of course.

Bare assertions by the source, "I disliked the movie," and "I consider Harry Truman one of the greater United States Presidents," have received considerable study as prestige suggestions. These may be considered inferences even though the inference may be a conditioned response of acceptance or rejection. Such statements have their weight determined by the source and their scale value by the semantic content of the assertion. With a design based on two prestige sources, Application 1 would provide a straightforward test of the integration model.

Consistency arguments may fall into both classes. "If you consider self-reliance a virtue, then you should allow your children more freedom" requires an assessment of the force of the implication. A means–end evaluation is also needed. Some consistency arguments are simple direct inferences, however.

This classification bridges the molar to the molecular. It is thus one approach to the detailed processing and integration of complex communications. Many communications will not stand a simple dissection, of course, but the total effect of those that do may be simple functions of the components. Suppose that $w_1, w_2, \cdots$, are the weights and $s_1, s_2, \cdots$, the scale values of the components of some communication. Averaging theory then implies that the scale value of the whole is the mean value of the parts,

$$s = \sum w_i s_i / \sum w_i, \qquad [21]$$

and that the weight of the whole is the sum of the weights of the parts,

$$w = \sum w_i. \qquad [22]$$

Equation 22 has interest as an adding rule within averaging theory. Inconsistency or redundancy among the components could affect their weights, of course. Theoretically, Equations 21 and 22 should still apply with these altered weights.

COGNITIVE CONSISTENCY

Cognitive consistency is central in attitude structure, and many current theories of attitude change take some consistency principle as their basic postulate and point of departure. The popularity of this approach is to be seen among the 84 chapters of *Theories of Cognitive Consistency: A Sourcebook* (Abelson, Aronson, McGuire, Newcomb, Rosenberg, & Tannenbaum, 1968). The variety, interest, and importance of the questions that have been raised by the consistency theorists are impressive.

But as the basis for a general theory of attitudes, a consistency postulate seems to be inherently inadequate. Much attitude change does not involve inconsistency resolution but only straightforward integration of information (Anderson, 1968a; Lindner, 1971). New information, including another's opinions and arguments, may alter attitudes and actions in the absence of any imbalance, incongruity, dissonance, or any other inconsistency. If this is correct, then a consistency principle is inherently too limited to support a general theory of attitudes. Just this limitation seems to be reflected in the conceptual insufficiencies of the various consistency principles, as well as in the problems that have arisen in attempts at quantitative treatment. Much of the difficulty seems to result from forcing the consistency principle into situations that do not involve inconsistency.

The present approach has developed differently, beginning with situations that do not require inconsistency resolution. The logic of this attack is straightforward. Inconsistency among the informational stimuli will affect their parameter values. The test of a simple averaging or adding model, though it requires some care, is straightforward if the stimulus parameters are constant. But if they vary across context, the analysis is markedly more difficult. It then becomes difficult to distinguish parameter changes from alternative integration rules. And it is also difficult to distinguish basic defects in the model from shortcomings in the experimental techniques.

The success of the averaging model with consistent information makes it possible to interpret deviations produced by integration of inconsistent information (Anderson & Jacobson, 1965). Moreover, success in the simpler situations helps validate the experimental techniques and the analytical procedures. The development of functional measurement theory, in particular, and the use of numerical ratings as equal-interval data have been central in this work. The support that this methodology has received in the simpler tasks provides part of the basis for attributing psychological meaning to discrepancies from the simple model in other situations (e.g., Lampel & Anderson, 1968; Oden & Anderson, 1971; Sidowski & Anderson, 1967; Slovic, 1969).

Inconsistency Resolution

Mechanisms for resolving inconsistency have been studied by numerous writers (e.g., Abelson, 1959, 1968b; Adams, 1968; Aronson, 1968; Feather, 1967; Festinger, 1957; Gollob, 1968; Hardyck & Kardush, 1968; Kaplan & Crockett, 1968; Kelman & Baron, 1968; McGuire, 1966, 1968; Weick, 1968). The present analysis has various similarities to previous treatments, though perhaps with some advantages in precision.

The most direct method of resolving inconsistency is in the valuation operation. If the informational stimuli are inconsistent, changes either in their meaning or in their importance could reduce the inconsistency. In the integration model, changes in meaning would be reflected as changes in s; changes in importance would be reflected as changes in w.

Attitude change experiments may explicitly or implicitly involve judgments of source and communication as well as the issue. The valuation process then becomes more complex as may be illustrated by considering an unpleasant communication attributed to a respected source. For the communication itself, the evidence below suggests that w will decrease but that s will remain constant. The decrease in w may result from revaluing the source or from inconsistency discounting, equivalent here to dissociating source and communication. At the same time, the communication constitutes a piece of information about the source. To the extent that it is not discounted, it will decrease the scale value of the source according to the analysis given in the section on balance theory.

Other mechanisms for resolving inconsistency will not be considered here, but it should be noted that the relation between valuation and integration needs closer analysis. The present discussion has implicitly viewed valuation as preceding integration. This is too simple a view: inconsistency can only exist as a consequence of an attempted integration. Valuation and integration must then proceed concurrently and a more molecular analysis is needed to delineate the temporal course of the total process.

Within the present framework, however, the two main modes of inconsistency resolution are change in meaning and change in weight. The evidence on these two modes will be discussed briefly to illustrate some of the problems.

Change of Meaning

The hypothesis that words change meaning as a function of context has great intuitive appeal. Certainly, it is true under certain conditions, as for words that have distinct alternative meanings (Anderson, 1968c). However, whether words with a single principal meaning change as a function of context is doubtful and certainly not well supported by existing data. Various lines of evidence have been considered, but none gives very convincing evidence for change of meaning in this case.

Evidence from primacy. One line of evidence that has received considerable study is the primacy effect in personality impressions first noted by Asch (1946). If good and bad traits are presented in serial order, the earlier traits have greater influence on the overall impression. Asch interpreted this in terms of change of meaning, as though the subject was selecting out those shades of meaning of the later words that would fit better with the earlier words (Anderson, 1965b; Asch, 1946). However, an extended series of experiments has provided little support for this view (see Anderson, 1965b, 1971a). Instead, it appears that the later adjectives keep a fixed scale value but get lower weight. Whether this results from discounting or attention decrement, however, is not yet completely clear (see Hendrick & Costantini, 1970).

In other work, Asch (1948) has argued that political statements change their meaning when attributed to different sources. These results, however, can be accounted for directly in terms of integration theory. It is only necessary to assume that the subject is attempting to judge the position of the source on the issue. For this he has two pieces of information: the statement itself and his prior opinion of the source's position. His judgment will be a composite or weighted average of these two informational stimuli and hence will vary directly with his opinion of the source. In this judgmental interpretation, it is not necessary to assume that the statement changes in any way.

Component judgments. Further support for a judgmental view comes from work on judgments of the single items of a combination. If subjects are asked to judge the likableness of each separate trait in a person description, there is a positive context effect: the rating of each trait shifts from its context-free value toward the values of the other traits (Anderson, 1966a; Anderson & Lampel, 1965). The obvious interpretation is that these judgments directly reflect change in meaning. Beyond plausibility, there is no satisfactory evidence for change of meaning, and the alternative hypothesis of a generalized halo effect is equally plausible. The available evidence (Anderson, 1971a; Kaplan, 1971) supports the judgmental view.

Contrast and assimilation. The concepts of contrast and assimilation have been widely but uncritically used in social psychology. Such effects, if real, would correspond to changes in meaning or scale value Assimilation, however, may result from composite judgmental processes as already suggested, and it is well known that contrast effects may be artifacts of the response language rather than changes in stimulus value. Even with psychophysical stimuli such as lifted weights, contrast effects are not always obtained (Anderson, 1971b; Schiffman, Goldstein, & Aroksaar, 1970).

The problem of contrast with verbal stimuli has been considered by numerous workers (e.g., Anderson & Lampel, 1965; Campbell, Lewis, & Hunt, 1958; Dawes, 1971; Manis, 1971; Parducci, 1965; Segall, 1959; Upshaw, 1965). It seems clear that most observed contrast effects stem largely from response language usage, not from true change in stimulus value. Dawes' (1971) report has special interest in this respect because it indicates that people with opposite views are perceived as more extreme than in fact they are. In addition, some evidence suggests that true stimulus contrast may be obtained under certain circumstances. Even if that is true, however, it does not justify the use of response measures that are severely contaminated by artifacts.

Judgmental context. The scale value of a word will certainly depend on the dimension of judgment. Masculinity is more desirable in men than women, for instance. This case is like that of words that have more than one principal meaning, and illustrates the role of the judgmental context in the valuation process. Similarly, the scale value of *lighthearted* would be different in judgments of likableness than of dependability.

Change of Weight

Discounting refers to decreased weighting owing to interaction among the informational stimuli. If a new piece of information is inconsistent with the old, the inconsistency can be reduced by assigning it less weight or importance. Direct evidence for inconsistency discounting in the personality-impression task is given by Anderson and Jacobson (1965) and Schümer and Cohen (1968). An important study using more realistic stimuli that bears on discounting processes has been made by Bugental, Kaswan, and Love (1970). Outlier discounting has also been studied (Anderson, 1968d). Moreover, the simplest form of redundancy interaction (e.g., Brewer, 1968; Dustin & Baldwin, 1966; Schmidt, 1969) would also correspond to decreases in weight. In contrast to the change of meaning hypothesis, then, the change of weight hypothesis seems to be on firm ground.

Stimulus interaction might possibly increase weight parameters. No evidence seems to exist, though increased weighting might occur when two stimuli form some natural unit. Chalmers (1969) has argued for both incremental and decremental weighting in personality impression formation by analogy with facilitation and interference in verbal learning. (It should be noted that Chalmers used the term, change of meaning, to refer to changes in the weight parameter.)

Discounting has received relatively little study, and that mainly with indirect measures, though an attempt to extend discounting analysis has been made by Schümer and Cohen (1968). Direct subjective estimates of importance would be most desirable, and some very tentative evidence for the usefulness of importance ratings is given in Anderson and Alexander (1971). Validation of such importance ratings would be especially valuable for configural analysis, since it is evident that discounting must depend on the configural pattern of the stimuli.

CONCLUDING COMMENTS

A distinctive feature of integration theory is its basis in a functional theory of measurement. Scaling thus becomes an organic part of the substantive investigation. The scales are developed and used directly in the substantive problem, and their validity rests on the validation of the psychological law. In some sense, this must be true of any measurement theory, but typically the scaling problem in psychology has been considered separately, to be accomplished as a methodological preliminary. In functional measurement theory, stimulus and response scaling and the psychological law are cofunctional in their development.

Integration theory is especially concerned with situations that require putting together several pieces of information. Such situations are not unique to opinions and attitudes. Learning, perception, judgment, and decision making also involve information integration. There should be a unified theory that covers all these areas. The present formulation has shown some initial promise in each of these areas, and it may provide a basis for the development of a unified general theory.

REFERENCES

ABELSON, R. P. Modes of resolution of belief dilemmas. *Journal of Conflict Resolution*, 1959, **3**, 343–352.

ABELSON, R. P. Computer simulation of "hot" cognition. In S. S. Tomkins & S. Messick (Eds.), *Computer simulation of personality*. New York: Wiley, 1963.

ABELSON, R. P. Mathematical models of the distribution of attitudes under controversy. In N. Frederiksen & H. Gulliksen (Eds.), *Contributions to mathematical psychology*. New York: Holt, Rinehart & Winston, 1964.

ABELSON, R. P. Psychological implication. In R. P. Abelson, E. Aronson, W. J. McGuire, T. M. Newcomb, M. J. Rosenberg, & P. H. Tannenbaum (Eds.), *Theories of cognitive consistency: A sourcebook*. Chicago: Rand McNally, 1968. (a)

ABELSON, R. P. A summary of hypotheses on modes of resolution. In R. P. Abelson, E. Aronson, W. J. McGuire, T. M. Newcomb, M. J. Rosenberg, & P. H. Tannenbaum (Eds.), *Theories of cognitive consistency: A sourcebook*. Chicago: Rand McNally, 1968. (b)

ABELSON, R. P., ARONSON, E., MCGUIRE, W. J., NEWCOMB, T. M., ROSENBERG, M. J., & TANNENBAUM, P. H. (Eds.) *Theories of cognitive consistency: A sourcebook*. Chicago: Rand McNally, 1968.

ADAMS, J. S. A framework for the study of modes of resolving inconsistency. In R. P. Abelson, E. Aronson, W. J. McGuire, T. M. Newcomb, M. J. Rosenberg, & P. H. Tannenbaum (Eds.), *Theories of cognitive consistency: A sourcebook*. Chicago: Rand McNally, 1968.

ANDERSON, L. R., & FISHBEIN, M. Prediction of attitude from the number, strength, and evaluative aspect of beliefs about the attitude object. *Journal of Personality and Social Psychology*, 1965, **2**, 437–443.

ANDERSON, N. H. Test of a model for opinion change. *Journal of Abnormal and Social Psychology*, 1959, **59**, 371–381. (a)

ANDERSON, N. H. An analysis of sequential dependencies. In R. R. Bush & W. K. Estes (Eds.), *Studies in mathematical learning theory*. Stanford: Stanford University Press, 1959. (b)

ANDERSON, N. H. Effect of first-order conditional probability in a two-choice learning situation. *Journal of Experimental Psychology*, 1960, **59**, 73–93.

ANDERSON, N. H. Scales and statistics: Parametric and nonparametric. *Psychological Bulletin*, 1961, **58**, 305–316. (a)

Anderson, N. H. Two learning models for responses measured on a continuous scale. *Psychometrika*, 1961, **26**, 391–403. (b)

Anderson, N. H. Application of an additive model to impression formation. *Science*, 1962, **138**, 817–818. (a)

Anderson, N. H. On the quantification of Miller's conflict theory. *Psychological Review*, 1962, **69**, 400–414. (b)

Anderson, N. H. Comparison of different populations: Resistance to extinction and transfer. *Psychological Review*, 1963, **70**, 162–179.

Anderson, N. H. Linear models for responses measured on a continuous scale. *Journal of Mathematical Psychology*, 1964, **1**, 121–142. (a)

Anderson, N. H. Note on weighted sum and linear operator models. *Psychonomic Science*, 1964, **1**, 189–190. (b)

Anderson, N. H. An evaluation of stimulus sampling theory: Comments on Prof. Estes' paper. In A. W. Melton (Ed.), *Categories of human learning*. New York: Academic Press, 1964. (c)

Anderson, N. H. Test of a model for number-averaging behavior. *Psychonomic Science*, 1964, **1**, 191–192. (d)

Anderson, N. H. Averaging versus adding as a stimulus-combination rule in impression formation. *Journal of Experimental Psychology*, 1965, **70**, 394–400. (a)

Anderson, N. H. Primacy effects in personality impression formation using a generalized order effect paradigm. *Journal of Personality and Social Psychology*, 1965, **2**, 1–9. (b)

Anderson, N. H. Component ratings in impression formation. *Psychonomic Science*, 1966, **6**, 279–280. (a)

Anderson, N. H. Test of a prediction of stimulus sampling theory in probability learning. *Journal of Experimental Psychology*, 1966, **71**, 499–510. (b)

Anderson, N. H. Averaging model analysis of set size effect in impression formation. *Journal of Experimental Psychology*, 1967, **75**, 158–165.

Anderson, N. H. A simple model for information integration. In R. P. Abelson, E. Aronson, W. J. McGuire, T. M. Newcomb, M. J. Rosenberg, & P. H. Tannenbaum (Eds.), *Theories of cognitive consistency: A sourcebook*. Chicago: Rand McNally, 1968. (a)

Anderson, N. H. Application of a linear-serial model to a personality-impression task using serial presentation. *Journal of Personality and Social Psychology*, 1968, **10**, 354–362. (b)

Anderson, N. H. Likableness ratings of 555 personality-trait words. *Journal of Personality and Social Psychology*, 1968, **9**, 272–279. (c)

Anderson, N. H. Averaging of space and number stimuli with simultaneous presentation. *Journal of Experimental Psychology*, 1968, **77**, 383–392. (d)

Anderson, N. H. Comment on "An analysis-of-variance model for the assessment of configural cue utilization in clinical judgment." *Psychological Bulletin*, 1969, **72**, 63–65. (a)

Anderson, N. H. Application of a model for numerical response to a probability learning situation. *Journal of Experimental Psychology*, 1969, **80**, 19–27. (b)

Anderson, N. H. Functional measurement and psychophysical judgment. *Psychological Review*, 1970, **77**, 153–170. (a)

Anderson, N. H. Averaging model applied to the size-weight illusion. *Perception and Psychophysics*, 1970, **8**, 1–4. (b)

Anderson, N. H. Two more tests against change of meaning in adjective combinations. *Journal of Verbal Learning and Verbal Behavior*, 1971, **10**, 75–85. (a)

Anderson, N. H. Test of adaptation-level theory as an explanation of a recency effect in psychophysical integration. *Journal of Experimental Psychology*, 1971, **87**, 57–63. (b)

Anderson, N. H., & Alexander, G. R. Choice test of the averaging hypothesis for information integration. *Cognitive Psychology*, 1971, in press.

Anderson, N. H., & Grant, D. A. A test of a statistical learning theory model for two-choice behavior with double stimulus events. *Journal of Experimental Psychology*, 1957, **54**, 305–317.

Anderson, N. H., & Grant. D. A. Correction and reanalysis. *Journal of Experimental Psychology*, 1958, **56**, 453–454.

Anderson, N. H., & Hovland, C. I. The representation of order effects in communication research. In C. I. Hovland (Ed.), *The order of presentation in persuasion*. New Haven: Yale University Press, 1957.

Anderson, N. H., & Hubert, S. Effects of concomitant verbal recall on order effects in personality impression formation. *Journal of Verbal Learning and Verbal Behavior*, 1963, **2**, 379–391.

Anderson, N. H., & Jacobson, A. Effect of stimulus inconsistency and discounting instructions in personality impression formation. *Journal of Personality and Social Psychology*, 1965, **2**, 531–539.

Anderson, N. H., & Jacobson, A. Further data on a weighted average model for judgment in a lifted weight task. *Perception and Psychophysics*, 1968, **4**, 81–84.

Anderson, N. H., & Lampel, A. K. Effect of context on ratings of personality traits. *Psychonomic Science*, 1965, **3**, 433–434.

Anderson, N. H., & Shanteau, J. C. Information integration in risky decision making. *Journal of Experimental Psychology*, 1970, **84**, 441–451.

Anderson, N. H., & Whalen, R. E. Likelihood judgments and sequential effects in a two-choice probability learning situation. *Journal of Experimental Psychology*, 1960, **60**, 111–120.

Aronson, E. Dissonance theory: Progress and problems. In R. P. Abelson, E. Aronson, W. J. McGuire, T. M. Newcomb, M. J. Rosenberg, & P. H. Tannenbaum (Eds.), *Theories of cognitive consistency: A sourcebook*. Chicago: Rand McNally, 1968.

Aronson, E., Turner, J., & Carlsmith, J. Communicator credibility and communicator discrepancy as determinants of opinion change. *Journal of Abnormal and Social Psychology*, 1963, **67**, 31–36.

Asch, S. E. Forming impressions of personality. *Journal of Abnormal and Social Psychology*, 1946, **41**, 258–290.

Asch, S. E. The doctrine of suggestion, prestige, and imitation in social psychology. *Psychological Review*, 1948, **55**, 250–276.

Atkinson, R. C., & Estes, W. K. Stimulus sampling theory. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology*. Vol. 2. New York: Wiley, 1963.

Bochner, S., & Insko, C. A. Communicator discrepancy, source credibility, and opinion change. *Journal of Personality and Social Psychology*, 1966, **4**, 614–621.

Bock, R. D., & Jones, L. V. *The measurement and prediction of judgment and choice*. San Francisco: Holden-Day, 1968.

Bogartz, R. S., & Wackwitz, J. H. Transforming response measures to remove interactions or other sources of variance. *Psychonomic Science*, 1970, **19**, 87–89.

Bogartz, R. S., & Wackwitz, J. H. Polynomial response scaling and functional measurement. *Journal of Mathematical Psychology*, 1971, in press.

Brewer, M. B. Averaging versus summation in composite ratings of complex social stimuli. *Journal of Personality and Social Psychology*, 1968, **8**, 20–26.

Brigham, J. C., & Cook, S. W. The influence of attitude on judgments of plausibility: A replication and extension. *Educational and Psychological Measurement*, 1970, **30**, 283–292.

Bugental, D. E., Kaswan, J. W., & Love, L. R. Perception of contradictory meanings conveyed by verbal and nonverbal channels. *Journal of Personality and Social Psychology*, 1970, **16**, 647–655.

Bush, R. R., & Mosteller, F. *Stochastic models for learning*. New York: Wiley, 1955.

Byrne, D. Attitudes and attraction. In L. Berkowitz (Ed.), *Advances in experimental social psychology*. Vol. 4. New York: Academic Press, 1969.

Byrne, D., Lamberth, J., Palmer, J., & London, O. Sequential effects as a function of explicit and implicit interpolated attraction responses. *Journal of Personality and Social Psychology*, 1969, **13**, 70–78.

Byrne, D., & Rhamey, R. Magnitude of positive and negative reinforcements as a determinant of attraction. *Journal of Personality and Social Psychology*, 1965, **2**, 884–889.

Campbell, D. T., Lewis, N. A., & Hunt, W. A. Context effects with judgmental language that is absolute, extensive and extra-experimentally anchored. *Journal of Experimental Psychology*, 1958, **55**, 220–228.

Cartwright, D., & Harary, F. Structural balance: A generalization of Heider's theory. *Psychological Review*, 1956, **63**, 277–293.

Chalmers, D. K. Meanings, impressions, and attitudes: A model of the evaluation process. *Psychological Review*, 1969, **76**, 450–460.

Coombs, C. H. *A theory of data*. New York: Wiley, 1964.

Coombs, C. H. Thurstone's measurement of social values revisited forty years later. *Journal of Personality and Social Psychology*, 1967, **6**, 85–91.

Darlington, R. B. Multiple regression in psychological research and practice. *Psychological Bulletin*, 1968, **69**, 161–182.

Dawes, R. M. An experimental analysis of the contrast effect and its implications for intergroup communication. *Journal of Personality and Social Psychology*, 1971, in press.

DuCharme, W. M. Response bias explanation of conservative human inference. *Journal of Experimental Psychology*, 1970, **85**, 66–74.

Dustin, D. S., & Baldwin, P. M. Redundancy in impression formation. *Journal of Personality and Social Psychology*, 1966, **3**, 500–506.

Edwards, W. Conservatism in human information processing. In B. Kleinmuntz (Ed.), *Formal representation of human judgment*. New York: Wiley, 1968.

Edwards, W., Lindman, H., & Savage, L. J. Bayesian statistical inference for psychological research. *Psychological Review*, 1963, **70**, 193–242.

Estes, W. K., & Burke, C. J. A theory of stimulus variability in learning. *Psychological Review*, 1953, **60**, 276–286.

Feather, N. T. A structural balance model of communication effects. *Psychological Review*, 1964, **71**, 291–313.

Feather, N. T. A structural balance approach to the analysis of communication effects. In L. Berkowitz (Ed.), *Advances in experimental social psychology*. Vol. 3. New York: Academic Press, 1967.

Festinger, L. *A theory of cognitive dissonance*. Stanford: Stanford University Press, 1957.

Fishbein, M., & Hunter, R. Summation versus balance in attitude organization and change. *Journal of Abnormal and Social Psychology*, 1964, **69**, 505–510.

Friedman, M. P., Carterette, E. C., & Anderson, N. H. Long-term probability learning with a random schedule of reinforcement. *Journal of Experimental Psychology*, 1968, **78**, 442–455.

Goldberg, L. R. Simple models or simple processes? Some research on clinical judgments. *American Psychologist*, 1968, **23**, 483–496.

Gollob, H. F. Impression formation and word combination in sentences. *Journal of Personality and Social Psychology*, 1968, **10**, 341–353.

Gulliksen, H. Measurement of subjective values. *Psychometrika*, 1956, **21**, 229–244.

Hammond, K. R., Hursch, C. J., & Todd, F. J. Analyzing the components of clinical inference. *Psychological Review*, 1964, **71**, 438–456.

Hardyck, J. A., & Kardush, M. A modest modish model for dissonance reduction. In R. P. Abelson, E. Aronson, W. J. McGuire, T. M. Newcomb, M. J. Rosenberg, & P. H. Tannenbaum

(Eds.), *Theories of cognitive consistency: A sourcebook.* Chicago: Rand McNally, 1968.

HEIDER, F. *The psychology of interpersonal relations.* New York: Wiley, 1958.

HENDRICK, C. Averaging versus summation in impression formation. *Perceptual and Motor Skills*, 1968, **27**, 1295–1302.

HENDRICK, C., & COSTANTINI, A. F. Effects of varying trait inconsistency and response requirements on the primacy effect in impression formation. *Journal of Personality and Social Psychology*, 1970, **15**, 158–164.

HIMMELFARB, S. Effects of cue validity differences in weighting information. *Journal of Mathematical Psychology*, 1970, **7**, 531–539.

HIMMELFARB, S., & SENN, D. J. Forming impressions of social class: Two tests of an averaging model. *Journal of Personality and Social Psychology*, 1969, **12**, 38–51.

HOFFMAN, P. J. The paramorphic representation of clinical judgment. *Psychological Bulletin*, 1960, **57**, 116–131.

HOVLAND, C. I. *The order of presentation in persuasion.* New Haven: Yale University Press, 1957.

HOVLAND, C. I., JANIS, I. L., & KELLEY, H. H. *Communication and persuasion.* New Haven: Yale University Press, 1953.

HULL, C. L. *Principles of behavior.* New York: Appleton-Century-Crofts, 1943.

INSKO, C. A. *Theories of attitude change.* New York: Appleton-Century-Crofts, 1967.

JONES, M. R. From probability learning to sequential processing: A critical appraisal. *Psychological Bulletin*, 1971, in press.

KAPLAN, B., & CROCKETT, W. H. Developmental analysis of modes of resolution. In R. P. Abelson, E. Aronson, W. J. McGuire, T. M. Newcomb, M. J. Rosenberg, & P. H. Tannenbaum (Eds.), *Theories of cognitive consistency: A sourcebook.* Chicago: Rand McNally, 1968.

KAPLAN, M. F. Context effects in impression formation: The weighted average versus the meaning change formulation. *Journal of Personality and Social Psychology*, 1971, in press.

KAPLAN, M. F., & OLCZAK, P. V. Attitude similarity and direct reinforcement as determinants of attraction. *Journal of Experimental Research on Personality*, 1970, **4**, 186–189.

KELMAN, H. C., & BARON, R. M. Determinants of modes of resolving inconsistency dilemmas: A functional analysis. In R. P. Abelson, E. Aronson, W. J. McGuire, T. M. Newcomb, M. J. Rosenberg, & P. H. Tannenbaum (Eds.), *Theories of cognitive consistency: A sourcebook.* Chicago: Rand McNally, 1968.

KIESLER, C. A., COLLINS, B. E., & MILLER, N. *Attitude change.* New York: Wiley, 1969.

KRANTZ, D. H., & TVERSKY, A. *Conjoint-measurement analysis of composition rules in psychology.* (Tech. Rep. MMPP 70-10) Ann Arbor: University of Michigan Mathematical Psychology Program, 1970.

KRUSKAL, J. B. Analysis of factorial experiments by estimating monotone transformations of the data. *Journal of the Royal Statistical Society* (*B*), 1965, **27**, 251–263.

LAMPEL, A. K., & ANDERSON, N. H. Combining visual and verbal information in an impression-formation task. *Journal of Personality and Social Psychology*, 1968, **9**, 1–6.

LEFFORD, A. The influence of emotional subject matter on logical reasoning. *Journal of General Psychology*, 1946, **34**, 127–151.

LEVIN, I. P., & SCHMIDT, C. F. A paired-comparisons paradigm for investigating person perceptions. *Psychonomic Science*, 1970, **20**, 359–361.

LINDNER, R. Associative and dissociative bands in personality impression formation. In R. G. Smith, *Speech-communication: Theory and models.* New York: Harper & Row, 1970.

LINDNER, R. Congruity, balance, and information integration in personality impressions. Unpublished doctoral dissertation, Indiana University, 1971.

LUCE, R. D., & TUKEY, J. W. Simultaneous conjoint measurement: A new type of fundamental measurement. *Journal of Mathematical Psychology*, 1964, **1**, 1–27.

MANIS, M. Context effects in communication: Determinants of verbal output and referential decoding. In M. H. Appley (Ed.), *Adaptation-level theory: A symposium.* New York: Academic Press, 1971.

MANIS, M., GLEASON, T. C., & DAWES, R. M. The evaluation of complex social stimuli. *Journal of Personality and Social Psychology*, 1966, **3**, 404–419.

MASSARO, D. W., & ANDERSON, N. H. Judgmental model of the Ebbinghaus illusion. *Journal of Experimental Psychology*, 1971, in press.

MCGUIRE, W. J. A syllogistic analysis of cognitive relationships. In M. J. Rosenberg et al. (Eds.), *Attitude organization and change.* New Haven: Yale University Press, 1960.

MCGUIRE, W. J. The current status of cognitive consistency theories. In S. Feldman (Ed.), *Cognitive consistency.* New York: Academic Press, 1966.

MCGUIRE, W. J. Theory of the structure of human thought. In R. P. Abelson, E. Aronson, W. J. McGuire, T. M. Newcomb, M. J. Rosenberg, & P. H. Tannenbaum (Eds.), *Theories of cognitive consistency: A sourcebook.* Chicago: Rand McNally, 1968.

MCGUIRE, W. J. The nature of attitudes and attitude change. In G. Lindzey & E. Aronson (Eds.), *The handbook of social psychology.* Vol. 3. (2nd ed.) Reading, Mass: Addison-Wesley, 1969.

MEEHL, P. E. *Clinical vs. statistical prediction.* Minneapolis: University of Minnesota Press, 1954.

MEEHL, P. E. The cognitive activity of the clinician. *American Psychologist*, 1960, **15**, 19–27.

NEWCOMB, T. M. Interpersonal balance. In R. P. Abelson, E. Aronson, W. J. McGuire, T. M. Newcomb, M. J. Rosenberg, & P. H. Tannenbaum

(Eds.), *Theories of cognitive consistency: A sourcebook.* Chicago: Rand McNally, 1968.

ODEN, G. C., & ANDERSON, N. H. Differential weighting in integration theory. *Journal of Experimental Psychology*, 1971, in press.

OSGOOD, C. E., & TANNENBAUM, P. H. The principle of congruity in the prediction of attitude change. *Psychological Review*, 1955, **62**, 42–55.

PARDUCCI, A. Category judgment: A range-frequency model. *Psychological Review*, 1965, **72**, 407–418.

PARDUCCI, A., THALER, L., & ANDERSON, N. H. Stimulus averaging and the context for judgment. *Perception and Psychophysics*, 1968, **3**, 145–150.

PEAK, H. Attitude and motivation. *Nebraska Symposium on Motivation*, 1955, **3**, 149–188.

PETERSON, C., & BEACH, L. R. Man as an intuitive statistician. *Psychological Bulletin*, 1967, **68**, 29–46.

RAMSAY, J. O., & CASE, B. Attitude measurement and the linear model. *Psychological Bulletin*, 1970, **74**, 185–192.

RHINE, R. J., & SEVERANCE, L. J. Ego-involvement, discrepancy, source credibility, and attitude change. *Journal of Personality and Social Psychology*, 1970, **16**, 175–190.

ROSENBAUM, M. E., & LEVIN, I. P. Impression formation as a function of source credibility and order of presentation of contradictory information. *Journal of Personality and Social Psychology*, 1968, **10**, 167–174.

ROSENBAUM, M. E., & LEVIN, I. P. Impression formation as a function of source credibility and the polarity of information. *Journal of Personality and Social Psychology*, 1969, **12**, 34–37.

ROSENBERG, M. J. Cognitive structure and attitudinal affect. *Journal of Abnormal and Social Psychology*, 1956, **53**, 367–372.

ROSENBERG, M. J. An analysis of affective-cognitive consistency. In M. J. Rosenberg et al. (Eds.), *Attitude organization and change.* New Haven: Yale University Press, 1960.

ROSENBERG, M. J. Discussion: Impression processing and the evaluation of old and new objects. In R. P. Abelson, E. Aronson, W. J. McGuire, T. M. Newcomb, M. J. Rosenberg, & P. H. Tannenbaum (Eds.), *Theories of cognitive consistency: A sourcebook.* Chicago: Rand McNally, 1968.

ROSENBERG, S. Mathematical models of social behavior. In G. Lindzey & E. Aronson (Eds.), *The handbook of social psychology.* Vol. 1. (2nd ed.) Reading, Mass: Addison-Wesley, 1968.

ROUANET, H., & ROSENBERG, S. Stochastic models for the response continuum in a determinate situation: Comparisons and extensions. *Journal of Mathematical Psychology*, 1964, **1**, 215–232.

SAVAGE, L. J. *The foundations of statistics.* New York: Wiley, 1957.

SAWYERS, B. K., & ANDERSON, N. H. Test of integration theory in attitude change. *Journal of Personality and Social Psychology*, 1971, in press.

SCHIFFMAN, H. R., GOLDSTEIN, J., & AROKSAAR, R. Response language and context effects. *Psychonomic Science*, 1970, **20**, 127–128.

SCHMIDT, C. F. Personality impression formation as a function of relatedness of information and length of set. *Journal of Personality and Social Psychology*, 1969, **12**, 6–11.

SCHÜMER, R., & COHEN, R. Eine Untersuchung zur sozialen Urteilsbildung. II. Bemerkungen zur verschiedenen konkurrierenden Modellen der Urteilsbildung. *Archiv für die gesamte Psychologie*, 1968, **120**, 180–202.

SEGALL, M. H. The effect of attitude and experience on judgments of controversial statements. *Journal of Abnormal and Social Psychology*, 1959, **58**, 61–68.

SHANTEAU, J. C. An additive model for sequential decision making. *Journal of Experimental Psychology*, 1970, **85**, 181–191. (a)

SHANTEAU, J. C. Component processes in risky decision judgments. Unpublished doctoral dissertation, University of California, San Diego, 1970. (b)

SHANTEAU, J. C., & ANDERSON, N. H. Test of a conflict model for preference judgment. *Journal of Mathematical Psychology*, 1969, **6**, 312–325.

SHERIF, M., & HOVLAND, C. I. *Social judgment: Assimilation and contrast effects in communication and attitude change.* New Haven: Yale University Press, 1961.

SHERIF, C. W., SHERIF, M., & NEBERGALL, R. E. *Attitude and attitude change.* Philadelphia: Saunders, 1965.

SIDOWSKI, J. B., & ANDERSON, N. H. Judgments of city-occupation combinations. *Psychonomic Science*, 1967, **7**, 279–280.

SLOVIC, P. Analyzing the expert judge: A descriptive study of a stockbroker's decision processes. *Journal of Applied Psychology*, 1969, **53**, 255–263.

SLOVIC, P., & LICHTENSTEIN, S. Comparison of Bayesian and regression approaches to the study of information processing in judgment. *Oregon Research Institute Monograph*, 1970, **10**(1).

SMITH, R. G. *Speech-communication: Theory and models.* New York: Harper & Row, 1970.

TANNENBAUM, P. H. The congruity principle revisited: Studies in the reduction, induction, and generalization of persuasion. In L. Berkowitz (Ed.), *Advances in experimental social psychology.* Vol. 3. New York: Academic Press, 1967.

TANNENBAUM, P. H. The congruity principle: Retrospective reflections and recent research. In R. P. Abelson, E. Aronson, W. J. McGuire, T. M. Newcomb, M. J. Rosenberg, & P. H. Tannenbaum (Eds.), *Theories of cognitive consistency: A sourcebook.* Chicago: Rand McNally, 1968.

TANNENBAUM, P. H., & NORRIS, E. L. Effects of combining congruity principle strategies for the reduction of persuasion. *Sociometry*, 1965, **28**, 145–157.

TESSER, A. Differential weighting and directed meaning as explanations of primacy in impression formation. *Psychonomic Science*, 1968, **11**, 299–300.

THURSTONE, L. L. The method of paired comparisons for social values. *Journal of Abnormal and Social Psychology*, 1927, **21**, 384–400.

TORGERSON, W. S. *Theory and methods of scaling.* New York: Wiley, 1958.

TRIANDIS, H. C., & FISHBEIN, M. Cognitive interaction in person perception. *Journal of Abnormal and Social Psychology*, 1963, **67**, 448–453.

TUCKER, L. R. Intra-individual and inter-individual multidimensionality. In H. Gulliksen & S. Messick (Eds.), *Psychological scaling: Theory and applications.* New York: Wiley, 1960.

TUKEY, J. W. Analyzing data: Sanctification or detective work? *American Psychologist*, 1969, **24**, 83–91.

UPSHAW, H. S. The effect of variable perspectives on judgments of opinion statements for Thurstone scales. *Journal of Personality and Social Psychology*, 1965, **2**, 60–69.

UPSHAW, H. S. The personal reference scale: An approach to social judgment. In L. Berkowitz (Ed.), *Advances in experimental social psychology.* Vol. 4. New York: Academic Press, 1969.

WEICK, K. E. The Panglossian word of self-justification. In R. P. Abelson, E. Aronson, W. J. McGuire, T. M. Newcomb, M. J. Rosenberg, & P. H. Tannenbaum (Eds.), *Theories of cognitive consistency: A sourcebook.* Chicago: Rand McNally, 1968.

WEISS, D. J., & ANDERSON, N. H. Subjective averaging of length with serial presentation. *Journal of Experimental Psychology*, 1969, **82**, 52–63.

WEISS, D. J., & ANDERSON, N. H. Use of rank order data in functional measurement. *Psychological Bulletin*, in press.

WIEST, W. M. A quantitative extension of Heider's theory of cognitive balance applied to interpersonal perception and self-esteem. *Psychological Monographs*, 1965, **79** (14, Whole No. 607).

WISHNER, J. Reanalysis of "Impressions of personality." *Psychological Review*, 1960, **67**, 96–112.

WYER, R. S. A quantitative comparison of three models of impression formation. *Journal of Experimental Research on Personality*, 1969, **4**, 29–41.

WYER, R. S. Quantitative prediction of belief and opinion change: A further test of a subjective probability model. *Journal of Personality and Social Psychology*, 1970, **16**, 559–570.

WYER, R. S., & DERMER, M. Effect of context and instructional set upon evaluations of personality-trait adjectives. *Journal of Personality and Social Psychology*, 1968, **9**, 7–14.

WYER, R. S., & GOLDBERG, L. A probabilistic analysis of the relationships among beliefs and attitudes. *Psychological Review*, 1970, **77**, 100–120.

YNTEMA, D. B., & TORGERSON, W. S. Man-computer cooperation in decisions requiring common sense. *IRE Transations*, 1961, HFE-2, 20–26.

ZAJONC, R. B. Cognitive theories in social psychology. In G. Lindzey & E. Aronson (Eds.), *The handbook of social psychology.* Vol. 1. (2nd ed.) Reading, Mass: Addison-Wesley, 1968.

ZINNES, J. L. Scaling. *Annual Review of Psychology*, 1969, **20**, 447–478.

(Received August 24, 1970)

Psychological Review
971, Vol. 78, No. 3, 207–228

A MODEL OF THE PERCEPTION OF SIMPLE GEOMETRIC FIGURES [1]

PAUL C. VITZ[2] AND THOMAS C. TODD

New York University

A model of how humans perceive straight-line two- and three-dimensional geometric figures, for example, triangles, trapezoids, cubes, is presented. The major characteristics of the model are: (*a*) a figure is represented as a hierarchy of perceptual elements—lines, angles, areas; and (*b*) the process of perceptually organizing a figure is portrayed as a stochastic sampling of the elements. In addition, the total number of trials to sample the elements corrected by a measure of redundancy is proposed as a measure of the figure's complexity. The model successfully predicts the judged complexity of a wide variety of two-dimensional figures; with some additional rationale it also predicts the perceived three-dimensionality of figures such as necker cubes which can be seen in either two or three dimensions. In general, the predictions of perceived three-dimensionality are supported but there are discrepancies and other important unresolved problems in the treatment of perceived three-dimensionality.

This paper outlines a model of how humans perceive simple geometric figures. By simple figures we mean two-dimensional representations of figures composed entirely of straight lines. Fimiliar examples are triangles, trapezoids, cubes, etc. The model has two major characteristics. First, a figure is represented as a hierarchy of perceptual elements; second, the process of perceptually organizing these elements is portrayed as a stochastic sampling process. In addition, the total number of sampling trials necessary for the complete sampling of a figure, corrected by a measure of the figure's redundancy, is proposed as a measure of the complexity of the figure.

THE MODEL: TWO-DIMENSIONAL FIGURES

Stimulus Elements

The stimulus elements for two-dimensional figures assumed by the model are lines, angles, and areas. Any simple geometric figure can be specified by enumerating and describing these elements, and the model assumes that the perception of a figure is also an evaluation of these elements. In general terms, these elements are the same as those of elementary geometry, although there are some minor modifications required by the kind of stimulus figures under consideration and by certain apparently overriding intuitive considerations.

Line. A line in our simple patterns is a straight line, except that it extends only to the intersection with another line and not beyond it. For example, in Figure 1 there are the following eight lines: AB, AC, AD, BC, BE, CD, CE, and DE. Our definition excludes the theoretically possible lines AE and BD.

Angle. An angle is created by the intersection of two lines. Angles are defined by the model only between lines which are adjacent within the figure. Again, in Figure 1, there are 12 angles: BAC, ABC, BCA, etc. The definition excludes angles

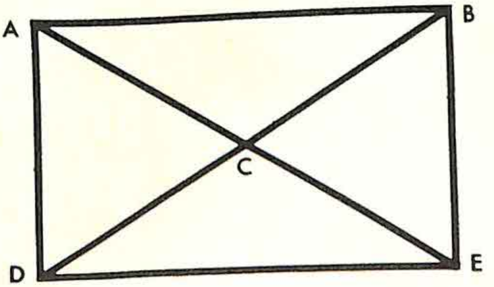


FIG. 1. Example of simple straight-line figure.

---

[1] This work was supported by Grant NS-07636-03 from the Institute of Neurological Diseases and Stroke, National Institutes of Health.

[2] Requests for reprints should be sent to Paul C. Vitz, Department of Psychology, New York University, 707 Broadway, New York, New York 10003.

such as BAD and ADE. In addition, an angle is defined as including the entire length of both the intersecting lines. This definition should be distinguished from definitions or uses of the term "angle" or "corner" which only include some unspecified length of each of the intersecting lines. Such uses, for example in Hebb's (1949, Ch. 5) theory of the elements involved in the perception of a triangle, have different implications for perception.

Area. Similarly, an area is restricted to the smallest component areas making up the entire figure. It will be seen by an examination of the example already given that these areas are generated by the lines and angles defined previously. In our example, the permissible areas are ABCA, ACDA, CDEC, and BCEB, but not ABECA or ABEDA.[3]

The major reason for the exclusion of the logically possible elements mentioned above is that the excluded elements would have overlapped with elements which are already included. That is, the included elements are the smallest set of nonoverlapping or independent elements which completely characterize the pattern.

Although the model's elements can be taken as assumptions, they do have a logical rationale. They are the simplest set of elements at each level. For example, the unbroken straight lines of a square are simpler than any other set of lines describing the square. We define the simpler of two sets of elements as that set with the fewer number of different elements; and if two sets have the same number of different elements, then the set with the least variance in element size is the simpler.

The reader might enquire why points, another basic geometric element, are not considered as elements for the model. Points falling on a line, although theoretically present, are assumed to have no perceptual existence. Points defined by the intersection of two lines are excluded, since whatever properties they have are assumed to be represented adequately by the lines and angles of which they are a part.

Sampling Process

The model represents the sampling of elements of a given kind as a process of independent sampling with replacement. For example, if a figure contains four lines of equal length, these lines are sampled with equal probabilities on each trial until the process stops sampling lines and begins sampling angles.

The model assumes that the probabilities of sampling of the elements at a given level are not necessarily the same, but instead are a function of the magnitudes of the elements. For example, consider a triangle with sides (lines) three, four, and five in length. The model assumes that the probability of sampling the side of Length 5 will be the greatest, and that for Length 3 the least. Specifically, the probability is equal to the proportion of the total magnitude contributed by that element. In the above example, the probability of sampling the line of Length 5 is $5/(3+4+5) = 5/12$; for Length 4, the probability is $4/12$; for Length 3, $3/12$. The magnitude or size of each of the classes of elements is defined as follows: of lines, it is the sum of the lengths of the lines; of angles, it is the sum of the interior angles; and of areas, it is the total area of the figure. Returning to our example of the triangle with sides of three, four, and five, we see that there are three lines, three angles, and one area. As we saw before, the line probabilities are $3/12$, $4/12$, and $5/12$. The angles of this triangle are 37°, 53°, and 90°, so that the corresponding sampling probabilities are $37/180$, $53/180$, and $90/180$. Finally, since there is only one area, it is sampled with probability one.

The perceptual process, then, is viewed as a process of sampling different classes of elements with probabilities assigned as outlined above. These classes of elements are arranged in a hierarchy, with the sampling process beginning with lines and ending with areas. All elements of a given class are sampled before the process moves on to the next class in the hierarchy.

---

[3] A fourth element, volumes, will be treated very briefly in the last section of this paper on three-dimensional perception.

Besides the mathematical justification for the preceding classes of elements, there is supporting or at least consistent empirical evidence. The well-known work of Hubel and Wiesel (1962, 1965) has demonstrated the existence of special bar or line detectors and special angle detectors in the visual cortex of the cat. Studies using stabilized images on the human retina suggest that lines, adjacent lines (angles), and planes (areas) all function as independent visual elements, since they seem to disappear and then reappear in an all-or-none fashion (Pritchard, Heron, & Hebb, 1960; Riggs & Tulunay, 1959). The recent work of Kaufman (Kaufman & Richards, 1969; Richards & Kaufman, 1969) has demonstrated that human subjects tend to fixate on the center of gravity of a geometric figure. This evidence is consistent with both the concept of an enclosed area as a higher order perceptual element and with the model's assumption, which is presented later, that perception involves a figure's symmetry often as evaluated about its center of gravity.

The model's assumption of sequential sampling of elements is consistent with the interpretation of perception proposed by McFarland (1965, 1968). McFarland presents evidence that human perception of a geometric figure consists of analysis or sequential part perception and then integration or the perception of the whole figure. Much of this research also implies a hierarchy of perceptual coding in which each level of the hierarchy is a temporally later and a more abstract stimulus property. Theoretical arguments for a hierarchically organized model of probabilistically sampled elements have been presented by Dodwell (1961). In addition, a recent model of the perception of sequential patterns by Vitz and Todd (1969) has demonstrated the usefulness of the stochastic sampling of hierarchically coded elements in representing the perceptual process.

The complexity of a figure unadjusted for presumed redundancies in the sampling process, which will be discussed later, is defined as the sum of the number of trials, $\sum T_i$, taken to sample the elements of each class. In the case of two-dimensional figures, the unadjusted complexity equals $\sum_{i=1}^{3} T_i$, when T_1 equals the trials to sample the lines, T_2 for angles, and T_3 for areas. The actual value of T_i for any particular set of probabilities is gotten by using a Monte Carlo method embodied within a simple computer program.[4] Thus the computer goes through the sampling process that the human visual system is portrayed as using. This sampling process must have a stop rule, since without it the process would go on indefinitely. The rationale for the stop rule is that the sampling stops after some fixed number of trials without sampling a new element. This number, symbolized SRC (stop rule criterion), is estimated from the data as described below. The perceptual interpretation of SRC is that the perceiver samples the elements of the figure until there has been a failure to detect a new element after SRC samples. Obviously the perceiver's certainty of having "seen" all the elements increases as the stop rule criterion increases. The major effect of SRC is to stop the sampling process before it spends inordinate numbers of trials "looking" for the last one or two extremely small elements. For this stop rule to work, it is also assumed that each element when sampled must be identified uniquely, and we assume this is done by tagging each element with its orientation and/or specific location. This information which is needed for the process to know if it has previously sampled an element is presumably related to the location of the sampled line in the population of line-sensing neural elements. In the case of K equally probable elements where $K \leqq 40$, T_i, with SRC$=75$, is an increasing approximately straight-line function of K; when SRC is smaller, the slope of the line is less.

There is the alternative of assuming the process knows when all elements have been sampled and then quits. Although this assumption is obviously unreasonable, it does allow one to represent the process by the expected number of trials, $E(T)$, for a fixed set of probabilities. In the case in which a class of elements is composed of K equally probable elements, the expected number of trials to sample this class, $E(T)$, is given by the following simple expression: $E(T) = K\ (1 + 1/2 + 1/3 + \cdots + 1/K)$ (Uspensky, 1937, p. 181).

---

[4] Copies of the computer program are available from the first author.

In the more common case where the elements are not equally probable, obtaining the expression for the expected number of trials requires some elementary matrix algebra and Markov chain theory.[5] The solution was developed and brought to our attention by Michael V. Levine and it is satisfactory when the number of elements, K, is small, since the number of terms in the matrix of the Markov chain for $E(T)$ is of the order of $2^k - 1$. For example, in the case of three elements a, b, and c with probabilities of p, q, and r, respectively, the total expected time before sampling all three, $E(T)$, is:

$$1+\frac{p}{1-p}+\frac{q}{1-q}+\frac{r}{1-r}+\frac{pq\quad(1+r)}{(1-p)(1-q)}+\frac{pr\quad(1+q)}{(1-p)(1-r)}+\frac{rq\quad(1+p)}{(1-r)(1-q)}.$$

Obviously when K is much larger than 5 or 6, the procedure is too cumbersome for hand computation.

We prefer the Monte Carlo method because it is fast, conceptually simple, based on a more plausible perceptual assumption, and because the stop rule parameter keeps the process from continuing for large numbers of trials just to pick up the last small element.

The procedure for estimating the size of SRC is simply to select two values—one high, one low, for example, 100 and 25—and move in by jumps of 5 or 10 until that value is found which gives the best fit between the predicted values and the data. In all experiments discussed here, the best estimation of SRC was 75. We expect this value of 75 will be typical for experiments in which the subject controls the length of time the stimulus is viewed before making a simple judgment. Of course the best fit SRC value would be smaller if the presentation were shortened tachistoscopically and/or if the experimenter introduced various costs associated with length of viewing time or with errors in reproducing or identifying the figure.

There is another question raised by the sampling process: How many sample runs are needed before the estimate T_i is assumed to be stable? That is, one sampling run to criterion can give a very unrepresentative estimate of T_i, especially for figures with many elements or with very discrepant probabilities. Throughout this paper, we used 3,000 runs for all T_i values, although it is likely several hundred runs would be sufficient, since with 3,000 runs the standard deviation of the T_i value was always less than 1%. In either case, we must assume that the perceptual system also makes large numbers of runs. Perhaps there are large numbers of independent subsets of line-sensing elements, each of which can be considered an "urn" and the number of trials or time to sample all or most of them can be considered as the analog to our estimate of T_i. That something like this is likely is argued by Ganz (1969). He discusses and presents physiological evidence that the visual cortex of the cat must have a great deal of redundancy since severe cortical loss still allows the animal to function effectively on visual discrimination tasks.

Examples of arrays of probabilities and the obtained Monte Carlo trials T_i to sample them are shown in Table 1.

The preceding definition of the sampling process assumed that each element was sampled independently of all other elements and with replacement. On intuitive grounds it seems extremely unlikely that such an assumption of complete independence is justified. For example, consider a square and a simple trapezoid—in Figure 4, No. 1 and 16. The number of independent trials to sample each pattern is about the same, since both figures have the same number of lines and angles, and the elements are not that different in size. This implies that

---

[5] The general rationale for finding $E(T)$ can be found in Kemeny and Snell (1960), especially Chapters II and III.

TABLE 1

EXAMPLES OF ARRAYS OF SAMPLING PROBABILITIES AND TRIALS TO SAMPLE, T_i, FOR THE LINES AND ANGLES OF STIMULUS FIGURES IN EXPERIMENT I

Stimulus numbers from Figure 4	Line probabilities	T_1	Angle probabilities	T_2
3—medium rectangle	(.1547, .1547, .3453, .3453)	10.52	$(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$	8.31
5—isosceles triangle	(.2062, .3969, .3969)	6.28	(.1666, .4167, .4167)	7.04
11—"stretched" hexagon	(.0933, .0933, .0933, .0933, .3134, .3134)	22.39	$(\frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6})$	14.75
16—simple trapezoid	(.1372, .2481, .2782, .3365)	9.99	(.1667, .2083, .2917, .3333)	9.38
19—"arrow"	(.1081, .1081, .1081, .1081, .2838, .2838)	19.41	(.0792, .0792, .1389, .1708, .1708, .3611)	21.48
22—"factory building"	(.0263, .0368, .0790, .0790, .1263, .1579, .2316, .2631)	43.53	(.0833, .0833, .0833, .0833, .0834, .0834, .2500, .2500)	29.47
25—"saw edge"	(.0410, .0410, .0410, .0525, .0525, .0702, .1170, .1170, .2047, .2631)	50.42	(.0312, .0416, .0416, .0416, .0625, .0625, .0950, .2080, .2080, .2080)	52.81

these figures are quite close in complexity. Intuitively, however, the square is considerably simpler, and its simplicity seems to reside in the fact that the sides and angles are identical. This property is closely related to the symmetry of the pattern. It seems very likely that identical elements can be sampled with greater efficiency, suggesting that some sort of redundancy be built into the representation of the sampling process. However, the present sampling process becomes unwieldly if dependencies representing element redundancies are incorporated directly in the sampling probabilities. Therefore, what is needed is a plausible rationale for some simple change in the representation of the sampling process that adjusts for pattern redundancy. This has been handled by the following rationale.

The general approach is to evaluate the symmetry of a figure on the assumption that a figure's degree of symmetry adequately reflects its redundancy. Symmetry is evaluated about a set of vertical and horizontal axes. (Other axes of symmetry are not treated.) Since it is assumed that symmetry is evaluated independently and in the same manner about each axis, it will suffice to present the procedure for evaluating vertical symmetry.

Definition 1: *Vertical axis of symmetry.* A figure's vertical axis of symmetry is a vertical line drawn through the figure's center of gravity which allows all elements, that is, all lines, angles, and areas of the figure, to map into each other about the axis. Common figures with this very simple kind of symmetry are shown in Figure 4, No. 1–5, 9–11, 13–15, 17, 20, and 21. Figures (in Figure 4) which also show horizontal symmetry are No. 1, 3, 4, 9, 11, 13, 15, 19, 20, and 21.

Definition 2a: Degree of symmetry, S, *for figures with vertical and horizontal symmetry.* A figure's degree of redundancy, S, is its number of axes of symmetry plus one. Thus, all two-dimensional figures with both a vertical and horizontal axis of symmetry have the maximum degree of redundancy of three.

Definition 3a: Two-dimensional figure complexity. The complexity of a symmetric two-dimensional figure is the sum of the number of trials to sample the lines, angles, and areas of the figure, $\sum_{i=1}^{3} T_i$, divided by S. This definition of complexity assumes that the redundancy in sampling a figure can be taken into account by the simple correction of dividing by the number of axes of symmetry plus one. (Adding one to the

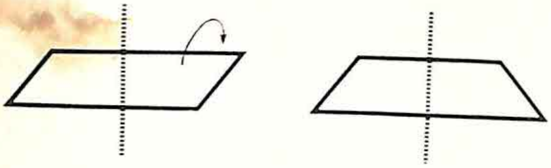


FIG. 2. Example of double mirror-image symmetry.

number of axes avoids dividing by zero in the case of completely asymmetric figures.) Thus, a figure with one axis of symmetry is divided by two and is assumed to be 50% redundant.

Definition 4: Vertical axis of double mirror-image symmetry. A vertical axis through the center of gravity of a figure creates double mirror-image symmetry if one of the two halves of the figure, when rotated about the horizontal axis, creates a new figure with simple or mirror-image symmetry. The parallelogram, Figure 2, is a familiar example of this kind of symmetry. Other examples of this kind of symmetry shown in Figure 4 are No. 6 (rhombus) and 23. In all cases of which we have any knowledge, figures with vertical double mirror-image symmetry also have horizontal double mirror-image symmetry.

Definition 2b: Degree of symmetry, S, *for figures with double mirror-image vertical and horizontal symmetry.* Each axis of double mirror-image symmetry counts as one half an axis of mirror-image symmetry. For example, a rhombus or parallelogram each have an S of $\frac{1}{2}+\frac{1}{2}+1$ or 2.00. The rationale for the choice of $\frac{1}{2}$ is that this kind of symmetry is obviously less than mirror-image symmetry. In addition two folds about the axis of symmetry are needed to get complete mapping, while only one fold is needed for mirror-image symmetry.

The two-dimensional complexity of a double mirror-image symmetric figure is the same as Definition 3, that is, $C=\sum T_i/S_i$.

Many figures which are not perfectly symmetric about one or both axes nevertheless have a great deal of redundancy which will be considered as a kind of symmetry called partial symmetry. The evaluation of partial symmetry is more involved, yet the basic idea is still closely related to simple symmetry. Again we wi consider only partial vertical symmetr since both axes are treated in the sam manner.

Definition 5: Partial vertical symmetry This symmetry is evaluated by placing vertical axis on the figure so as to maximiz the proportion of corresponding element which map into each other, or, if no elements map, then the vertical axis is place so as to maximize the proportion of corresponding elements which are identical with respect to their size or orientation.

Definition 6: Corresponding elements. Elements on either side of a vertical axis are corresponding if they are the same number of elements in distance from the same point of intersection of the vertical axis with the figure. If the distance of an element can be measured from more than one point of intersection of the vertical axis, then it is measured only from the nearest point of intersection. If an element should have more than one corresponding element, then it is tested against all corresponding elements until one, but no more than one, identity of size or orientation is found, if there is any. Corresponding elements are shown in Figure 3.

Definition 7: Point of intersection. Any point where the vertical or horizontal axis intersects the figure is a point of intersection. Examples are marked by arrows in Figure 3.

Definition 8: The size of an element. This was defined above in the discussion of the sampling probabilities.

Definition 9a: The orientation of a line is the angle of intersection of the line with the vertical axis.

Definition 9b: The orientation of an angle is the orientation of the line which bisects the angle.

Definition 9c: The orientation of an area is the orientation of the area considered as a plane in the third dimension. For all two-dimensional figures, the orientation of areas are the same. In the case of two-dimensional figures with more than one area, this causes a small problem. In such cases it is assumed that the correspondence of two-dimensional areas is evaluated on

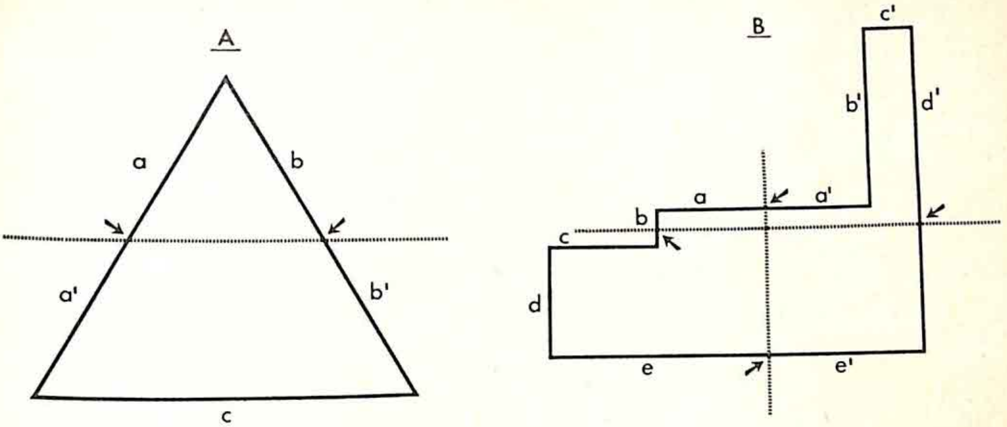


FIG. 3. Examples of partial symmetry. (Dotted lines are axes of partial symmetry; identical letters, for example, a and a′, designate corresponding lines; arrows denote points of intersection with axis.)

the dimension of size and also shape. Adding shape provides two characteristics as the evaluation of all other elements and allows only perfectly symmetric figures to have maximum S values of 3.

Examples of partial symmetry with special reference to corresponding elements are shown in Figure 3.

Definition 10: Degree of correspondence. On the dimensions of element size, the degree of correspondence is the proportion of corresponding elements that are identical in size. The degree of correspondence on the dimension of element orientation is the proportion of corresponding elements identical in orientation.

Definition 2c: Degree of symmetry, S, *for figures with partial symmetry.* In this case, S is not necessarily the same for both vertical and horizontal symmetry; nor is S necessarily the same for each class of elements, for example, the redundancy of its angles. For this latter reason, S is evaluated for each class of elements separately in order to correct the trials to sample each class by the appropriate redundancy estimate. The S for the lines is symbolized S_1; S for the angles S_2; etc. For lines, S_1 equals the degree of correspondence of the figure's lines about the vertical axis plus the degree of correspondence of the lines about the horizontal axis plus one; S_2 is defined analogously for the angles.

Definition 3b. The two-dimensional complexity of a figure with partial symmetry is $\sum_{i=1}^{3} (T_i/S_i)$. The slight difference between this definition and 3a is that S is subscripted, symbolizing the measure of redundancy specific of the T_i for each class of elements. Because it is more general, $\sum_{i=1}^{n} (T_i/S_i)$ typically will be used as the formula for complexity.

Table 2 presents degree of correspondence and S_i for the partially symmetric figures shown in Figure 4. A little study of it along with the examples discussed below will quickly familiarize the reader with this part of the model.

Two examples of the computation of S for figures with partial symmetry will be given. Figure 3A has perfect vertical symmetry; its partial horizontal symmetry is evaluated about the dotted horizontal axis. This axis has been set up so as to bisect the two nonhorizontal sides. These bisected or split elements are special cases with a special rationale for their treatment in computing S. Obviously the two parts of a split line always have the same orientation and they are always corresponding elements. This artifact should not contribute to a measure of figure redundancy, and therefore the model does not include split elements in estimating S *unless* there is another line, or element, also split by the same axis that has the same orientation. This is not the case in Figure 3A where the two split lines have different orientations. If they had had the same orientation, then the lines would have been treated as $\frac{2}{3}$ redundant on the dimension of orientation. In addition, an axis can always be put through a line so as to bisect it and thus to automatically set up two parts that correspond and are identical in length. Again in evaluating split elements, the model ignores this artifactual matching *unless* the same axis also bisects another line or element. In Figure 3A this is the case; thus for lines on the dimension of size, the figure is $\frac{2}{3}$ redundant and its total horizontal line redundancy, S_{1H}, is $(.00 + .67)/2$ or .33, and therefore the T_1 to sample the lines of Figure 3A is divided by the S_1 of 1.00 (vertical symmetry) + .33 (horizontal symmetry) + 1.00, which equals 2.33 (see Table 2).

In Figure 3B, both axes for evaluating partial symmetry are shown. None of Figure 3B's eight lines are of the same size when matched with their corresponding elements about either axis. (Note: The split elements do not count here since no axis bisects two lines.) However, all of the lines match their corresponding line in orientation about both axes. (Note: The split elements count since both lines split by each axis have the same orientation.) Thus, both S_{1V} and $S_{1H} = (.00 + 1.00)/2 = .50$; and $S_1 = .50 + .50 + 1.00 = 2.00$ (see Table 2).

When evaluated about the vertical axes on size, four of the angles of Figure 3B are identical in size, and four are identical in orientation. The same is true when evaluated about the horizontal axis. Thus S_{2V} and S_{2H} both equal $(.50 + .50)/2$ and

TABLE 2

DEGREE OF CORRESPONDENCE AND DEGREE OF SYMMETRY, *S*, FOR THE LINES AND ANGLES OF STIMULUS FIGURES WITH PARTIAL SYMMETRY USED IN EXPERIMENT I

Stimulus number from Figure 4	Lines					Angles				
	A % vertical correspondence (partial symmetry)		B % horizontal correspondence (partial symmetry)		S_1 $\frac{A}{2}+\frac{B}{2}+1$	C % vertical correspondence (partial symmetry)		D % horizontal correspondence (partial symmetry)		S_2 $\frac{C}{2}+\frac{D}{2}+1$
	Size	Orientation	Size	Orientation		Size	Orientation	Size	Orientation	
2	100 (maps)	100 (maps)	67	0	2.33	100 (maps)	100 (maps)	67	0	2.33
5	100 (maps)	100 (maps)	67	0	2.33	100 (maps)	100 (maps)	0	0	2.00
7	0	0	67	0	1.33	0	0	0	0	1.00
10	100 (maps)	100 (maps)	80	0	2.40	100 (maps)	100 (maps)	80	0	2.40
12	0	0	67	0	1.33	0	0	0	0	1.00
14	100 (maps)	100 (maps)	86	0	2.43	100 (maps)	100 (maps)	86	0	2.43
16	0	50	50	50	1.75	0	0	0	0	1.00
17	100 (maps)	100 (maps)	75	100	2.87	100 (maps)	100 (maps)	50	50	2.50
18	0	50	50	50	1.75	0	0	0	0	1.00
19	100	100	100 (maps)	100 (maps)	3.00	0	33	100 (maps)	100 (maps)	2.16
22	0	100	0	100	2.00	50	50	50	50	2.00
24	0	0	0	0	1.00	0	0	0	0	1.00
25	80	100	50	80	2.55	20	60	0	60	1.70
26	0	0	0	0	1.00	0	0	0	0	1.00
27	0	0	0	0	1.00	0	0	0	0	1.00
28	25	25	25	50	1.62	0	0	0	0	1.00

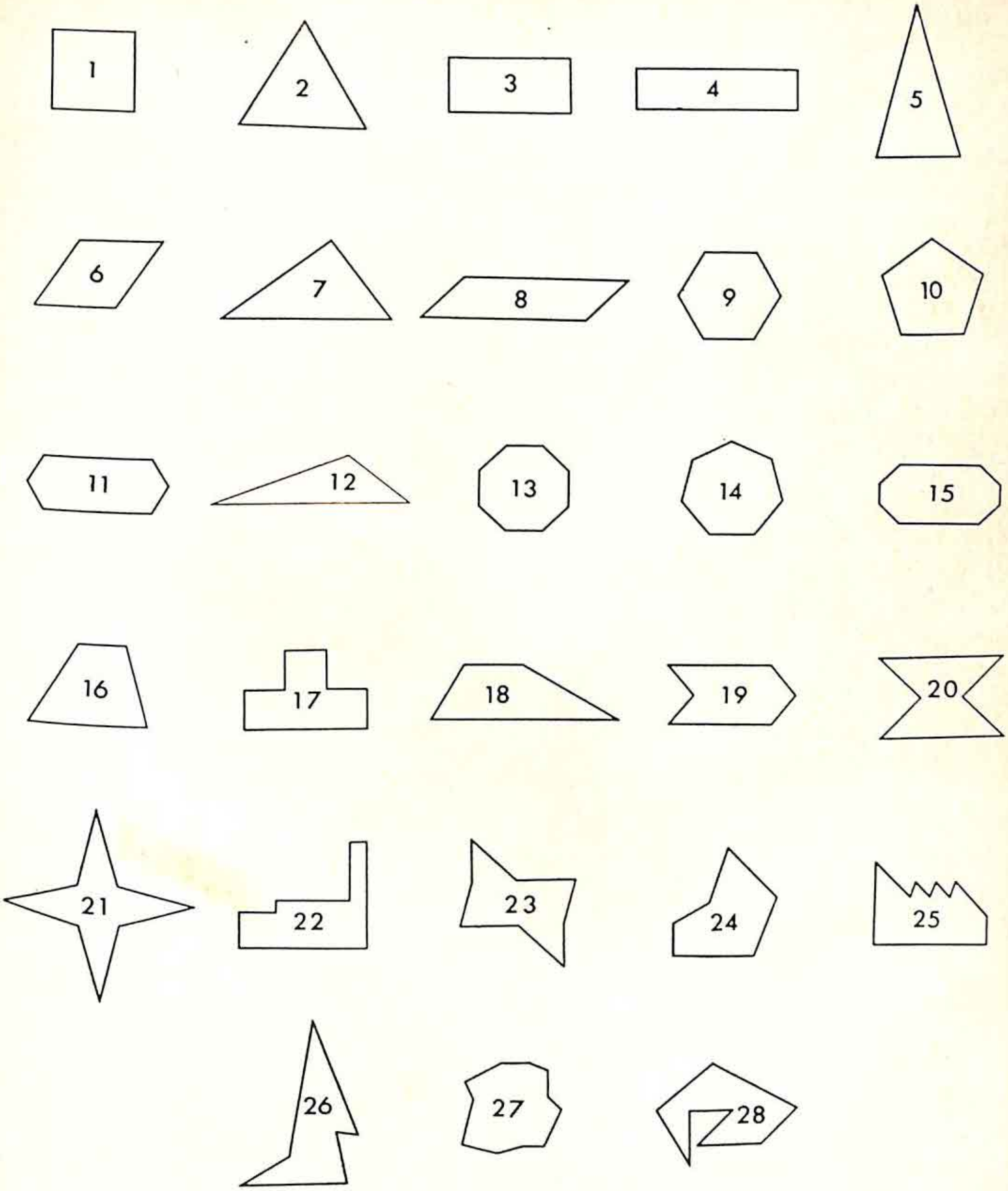


FIG. 4. The stimulus figures used in Experiment I. (Figures are in order of increasing judged complexity.)

$S_2 = .50 + .50 + 1.00 = 2.00$ (see Table 2). In checking these computations, the reader should recall that angle orientation is the orientation of the line bisecting the angle.

In the case of two-dimensional figures with only one area, the model does not define S. In such cases $T_3 = 1.00$.

The preceding discussion of the redundancy evaluation for partially symmetric figures may seem complex and/or difficult. In fact, with a little practice it becomes quite routine. In addition, it is conceptually a straightforward extension of the notion of perfect mapping of elements to figures where some of the elements map on only one of the relevant dimensions.

The formal presentation of the model may have disguised some of its simple fundamental characteristics. First the complexity of a figure is determined by three basic properties (everything else equal): (*a*) the more elements in a figure, the greater its complexity; (*b*) the more the elements differ in size, the greater its complexity; and (*c*) the greater the symmetry, the less the total complexity. These three properties can often allow one to decide which of two figures the model will predict as more complex prior to actual computation of the T values.

According to the model, the simplest closed geometric figure (see Table 4) is the equilateral triangle. This receives some support in the work of Hanes (1950), Bitterman, Krauskopf, and Hochberg (1954), and Borg (1964).

In terms of the traditional Gestalt and S–R interpretations of figure perception, this model is something of a mixture. In an important sense, the model does claim that perception is equal to the sum of its parts. Both the parts and the process of summing them are central to the model. On the other hand, the selection of simple, coded elements and their organization in a hierarchy has strong similarities to the Gestalt emphasis and is quite different from the normal S–R or elementist position.

SUPPORTING DATA: TWO-DIMENSIONAL FIGURES

In specifying the two-dimensional complexity of any simple geometric figure, the model predicts judgments of the complexity of different figures. Experiment I was designed to test this property of the model.

TABLE 3

JUDGED COMPLEXITY (MEAN RANK) AND BREAKDOWN OF THE COMPLEXITY EVALUATION OF THE STIMULUS FIGURES USED IN EXPERIMENT I

Stimulus number from Figure 4	Lines			Angles			Total figure complexity[a]	Judged complexity mean rank ($N = 20$)
	T_1	S_1	C_1	T_2	S_2	C_2	$\sum_{i=1}^{2} T_i/S_i$	
1	8.31	3.00	2.77	8.31	3.00	2.77	5.54	2.00
2	5.48	2.33	2.35	5.48	2.33	2.35	4.70	2.85
3	10.52	3.00	3.47	8.31	3.00	2.77	6.24	3.50
4	15.32	3.00	5.11	8.31	3.00	2.77	7.88	4.30
5	6.28	2.33	2.70	7.04	2.00	3.52	6.22	4.75
6	8.31	2.00	4.16	9.89	2.00	4.95	9.11	8.35
7	5.94	1.33	4.46	6.67	1.00	6.67	11.13	8.95
8	11.98	2.00	5.99	12.50	2.00	6.25	12.24	9.15
9	14.75	3.00	4.92	14.75	3.00	4.92	9.84	11.40
10	11.36	2.40	4.73	11.36	2.40	4.73	9.46	12.10
11	22.39	3.00	7.46	14.75	3.00	4.92	12.38	12.50
12	6.71	1.33	5.05	10.44	1.00	10.44	15.49	12.85
13	21.90	3.00	7.30	21.90	3.00	7.30	14.60	13.15
14	18.05	2.43	7.43	18.05	2.43	7.43	14.86	13.75
15	30.98	3.00	10.33	21.90	3.00	7.30	17.63	14.50
16	9.99	1.75	5.71	9.38	1.00	9.38	15.09	14.70
17	25.60	2.87	8.92	29.04	2.50	11.62	20.54	15.60
18	11.14	1.75	6.37	14.50	1.00	14.50	20.87	15.95
19	19.41	3.00	6.47	21.49	2.16	9.95	16.42	17.75
20	17.39	3.00	5.80	32.11	3.00	10.70	16.50	18.15
21	21.90	3.00	7.30	60.81	3.00	20.27	27.57	20.00
22	43.54	2.00	21.77	29.47	2.00	14.74	36.51	20.35
23	22.38	2.00	11.19	37.45	2.00	18.73	29.92	22.10
24	17.33	1.00	17.33	17.96	1.00	17.96	35.29	22.90
25	50.42	2.55	19.77	52.81	1.70	31.06	50.83	25.05
26	25.72	1.00	25.72	35.19	1.00	35.19	60.91	25.35
27	38.60	1.00	38.60	34.91	1.00	34.91	75.51	26.50
28	23.29	1.62	14.38	42.43	1.00	42.43	56.81	27.35

[a] The contribution due to the single area of each figure has been omitted since it is negligible and constant for each figure. The ame procedure was followed in Experiment II (Table 4).

Experiment I

Method

Subjects. The subjects were 20 adults, 10 male and 10 female, with a median age of 22 and an age range 17–48; they were undergraduate and graduate students, secretaries, and technicians.

Stimuli. The 28 geometric figures used as stimuli are shown in Figure 4. These particular figures were selected on the general intuitive grounds that they represented a wide variety of different features. All figures had the same area and were thus of approximately equal perceived size. Each figure was photographed on glossy paper, cut out inside of a 7-inch circle and covered with clear plastic to protect the photograph and to aid in handling. The figures were oriented as shown in Figure 4. Orientation was designated by a small dot at the bottom of the card.

Procedure. After looking through the deck of 28 stimulus disks, the subject was asked to divide the stimuli into two piles: the 14 stimuli he judged to be the simplest and the 14 he judged to be more complex. The subject was allowed to use any definition of simple or complex he wished. The subject was asked to rank the 14 simpler from simplest (Rank 1) to most complex (Rank 14). The first 10 of these ranks were recorded, and the corresponding stimuli permanently set aside. The remaining 4 stimuli were mixed in with the original 14 more complex stimuli, and from these 18 stimuli, the subject was asked to temporarily put aside the four most complex stimuli. The remaining 14 stimuli were then ranked for simplest (Rank 11) to most complex (Rank 24). Ranks 11–20 were recorded and the corresponding stimuli permanently set aside. The remaining four stimuli plus the four set aside as most complex (Ranks 21–28) were then ranked. This modified ranking procedure, taking about 15 to 20 minutes, was not difficult, and by breaking down the task into smaller sets, aided the subject in his rankings.

Results

The major results are shown in Table 3 and in Figure 5. Table 3 presents a summary of the complexity of the lines and angles of each of the 28 figures shown in Figure 4. Table 3 also shows the total complexity of each figure based on a stop rule criterion of 75 and the obtained judged complexity (mean rank). Figure 5 is a graph of the relation between the total (predicted) and the judged (obtained) complexity values. To get a quantitative estimate of the relation shown in Figure 5, we computed a product-moment correlation between $\log_{10}$ of the predicted complexity value and the judged complexity. This correlation was .97 and obviously represents considerable support for the model.

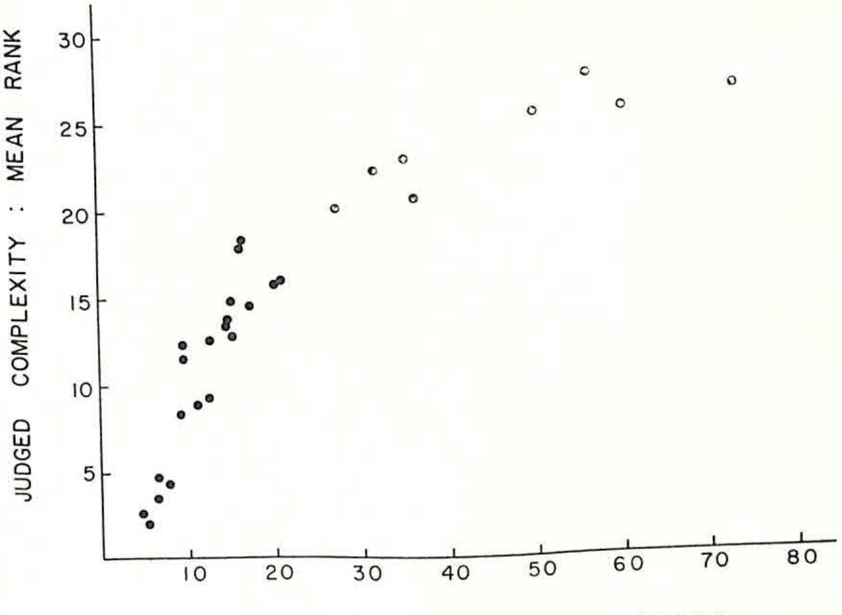


FIG. 5. Results of Experiment I. (Judged complexity [mean rank, $N = 20$] as a function of predicted figure complexity.)

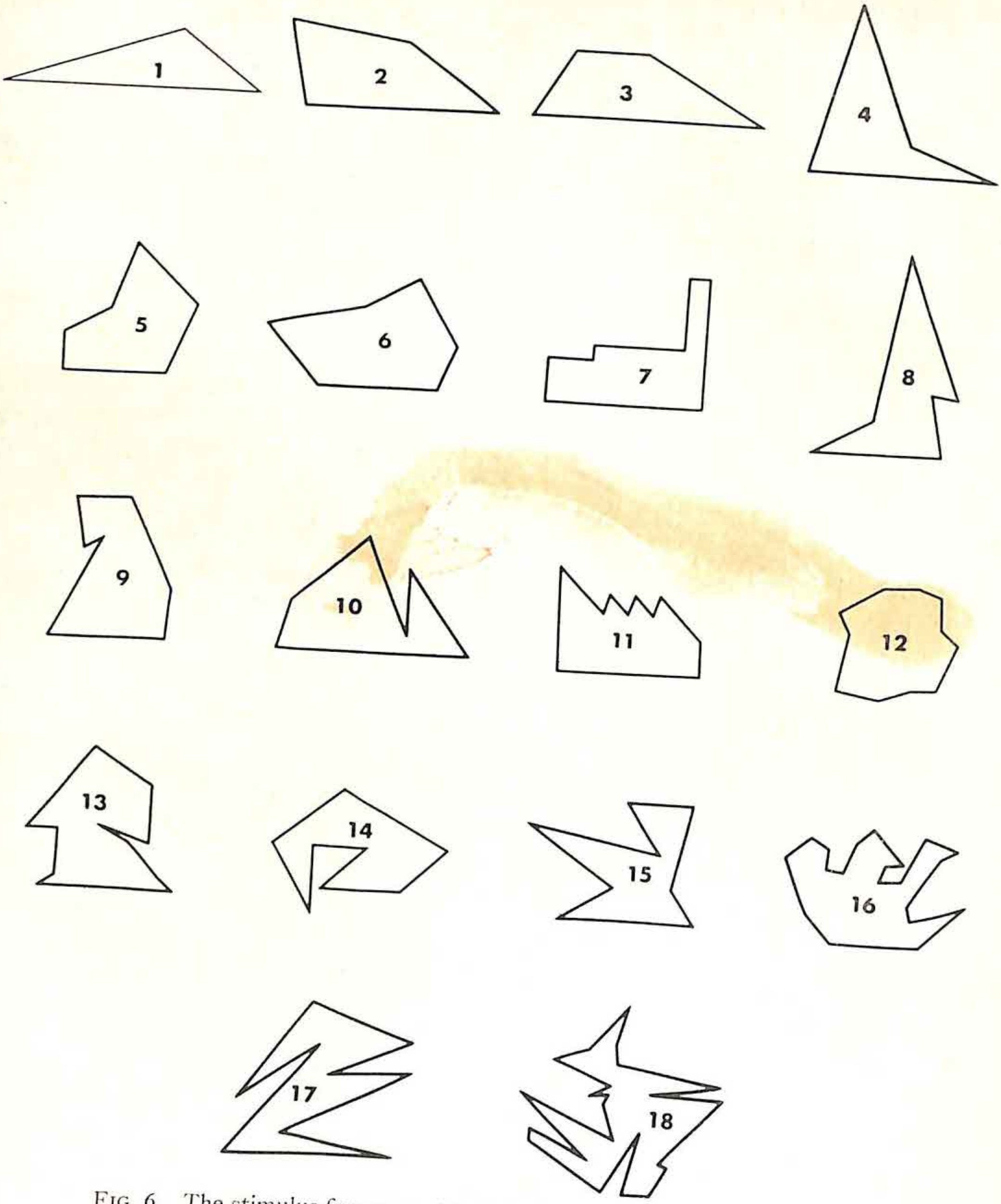


FIG. 6. The stimulus figures used in Experiment II. (Figures are in order of increasing judged complexity.)

A possible correction in the model may be implied by a small discrepancy in the results of Experiment I. In Figure 4, No. 22, which consists of entirely vertical and horizontal lines, seems to have a predicted value that is slightly too large (see Table 3). There is considerable evidence that humans perceive vertical and horizontal lines with greater efficiency than lines at other orientations (Campbell, Kulikowski, & Levinson, 1966; Higgins & Stultz, 1948). The model assumes all orientations are sampled with equal probability and this may need correction.

Experiment II

A large number of the figures used in Experiment I were highly symmetric. In

Experiment II, 18 primarily asymmetric figures were used to test again the model's capacity to predict judgments of figure complexity.

Method

Subjects. The 20 subjects, 10 males and 10 females, ranged in age from 12 to 70; median age 25. They were selected to get a variety of ages and professional backgrounds.

Stimuli. The stimulus figures are shown in Figure 6. Eight of these figures, No. 1, 3, 5, 7, 8, 11, 12, and 14, are No. 12, 18, 24, 22, 26, 25, 27, and 28, respectively, in Experiment I. The remaining 10 figures, No. 2, 4, 6, 9, 10, 13, 15, 16, 17, and 18, are Vanderplas shapes—their Vanderplas numbers are as follows: Figure 2, No. 9 and 25; Figure 3, No. 11 and 28; Figure 4, No. 5 and 11; Figure 5, No. 15 and 17; Figure 7, No. 2 and 12 (see Vanderplas & Garvin, 1959, pp. 149–151). The figures were constructed with equal areas, photographed, cut out within a circle approximately 7 inches in diameter, and covered with clear plastic. The orientation, shown in Figure 6, was designated by a dot on the plastic cover.

Procedure. A modified ranking procedure similar to Experiment I was used. The subjects were given the set of 18 figures, in a scrambled order; they were told to divide them into two sets—the nine most simple and the nine most complex. They could use any definitions of simple or complex they wished. Next they ranked the nine simplest shapes in order from simplest to most complex. The first six figures (ranks) were recorded; the other three were returned to the set of the nine most complex figures. From this set of 12 figures, the subject removed the three he judged as most complex. The remaining nine figures were ranked from simplest to most complex. The first six of these figures (ranks) were recorded. The remaining three figures plus the three judged as most complex formed the final set of six which were then ranked. The procedure lasted about 15–20 minutes.

Results

The major results are shown in Table 4 and Figure 7. Figure 7 shows that in general the model predicts the judged complexity values quite well. As in Experiment I, $\log_{10}C$ values were used to correct for the ceiling effect in the judgments. The product-moment correlation between $\log_{10}C$ and the judged complexity rank is .96.

Brown's Random Polygons

One difficulty with the previous experiments is that the representativeness or

TABLE 4

T_1, T_2 FOR THE VANDERPLAS FIGURES USED IN EXPERIMENT II; TOTAL FIGURE COMPLEXITY, LOG COMPLEXITY, AND JUDGED COMPLEXITY FOR ALL FIGURES USED IN EXPERIMENT II

Stimulus number from Figure 6	Lines T_1	Angles T_2	Total figure complexity $C = \sum_{i=1}^{2} T_i/S_i$ $S = 1.00$ for all Vanderplas figures	$\text{Log}_{10}\, C$	Judged complexity mean rank ($N = 20$)
1 (12)[a]	[a]	[a]	15.49	1.190	1.05
2	9.25	11.30	20.55	1.313	2.70
3 (18)[a]	[a]	[a]	20.87	1.319	2.90
4	9.29	21.18	30.47	1.484	5.00
5 (24)[a]	[a]	[a]	35.29	1.548	5.40
6	16.57	17.54	34.11	1.533	6.50
7 (22)[a]	[a]	[a]	36.51	1.562	6.85
8 (26)[a]	[a]	[a]	60.91	1.785	9.25
9	27.79	26.15	53.94	1.732	9.40
10	18.96	28.13	47.09	1.673	9.85
11 (25)[a]	[a]	[a]	50.83	1.706	11.35
12 (27)[a]	[a]	[a]	73.51	1.866	12.25
13	49.90	61.51	111.41	2.047	12.65
14 (28)[a]	[a]	[a]	56.81	1.754	12.70
15	26.89	52.51	79.40	1.900	13.35
16	95.07	98.17	193.24	2.286	16.00
17	27.55	74.37	101.92	2.008	16.20
18	119.76	141.25	261.01	2.417	17.95

[a] See Table 3 for these numbers and complexity values.

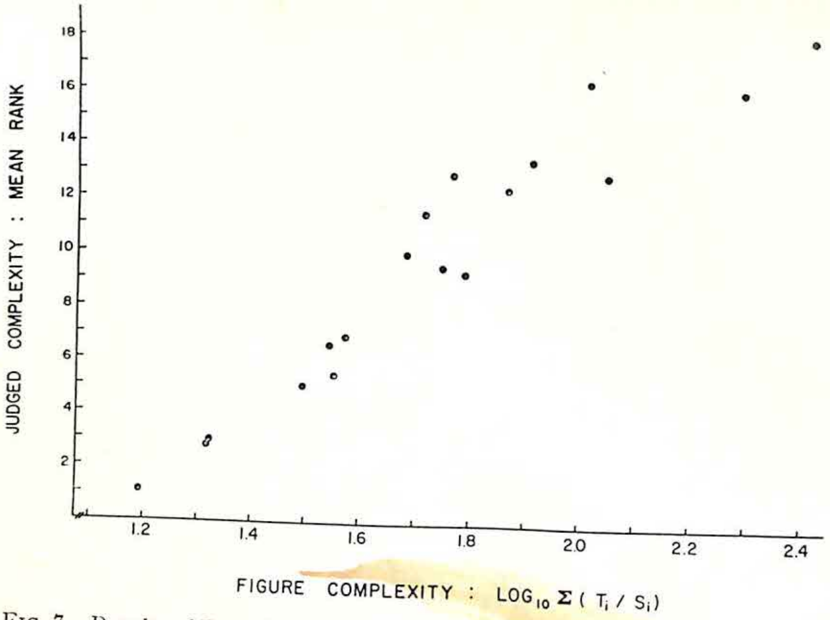


FIG. 7. Results of Experiment II. (Judged complexity [mean rank, $N = 20$] as a function of predicted figure complexity.)

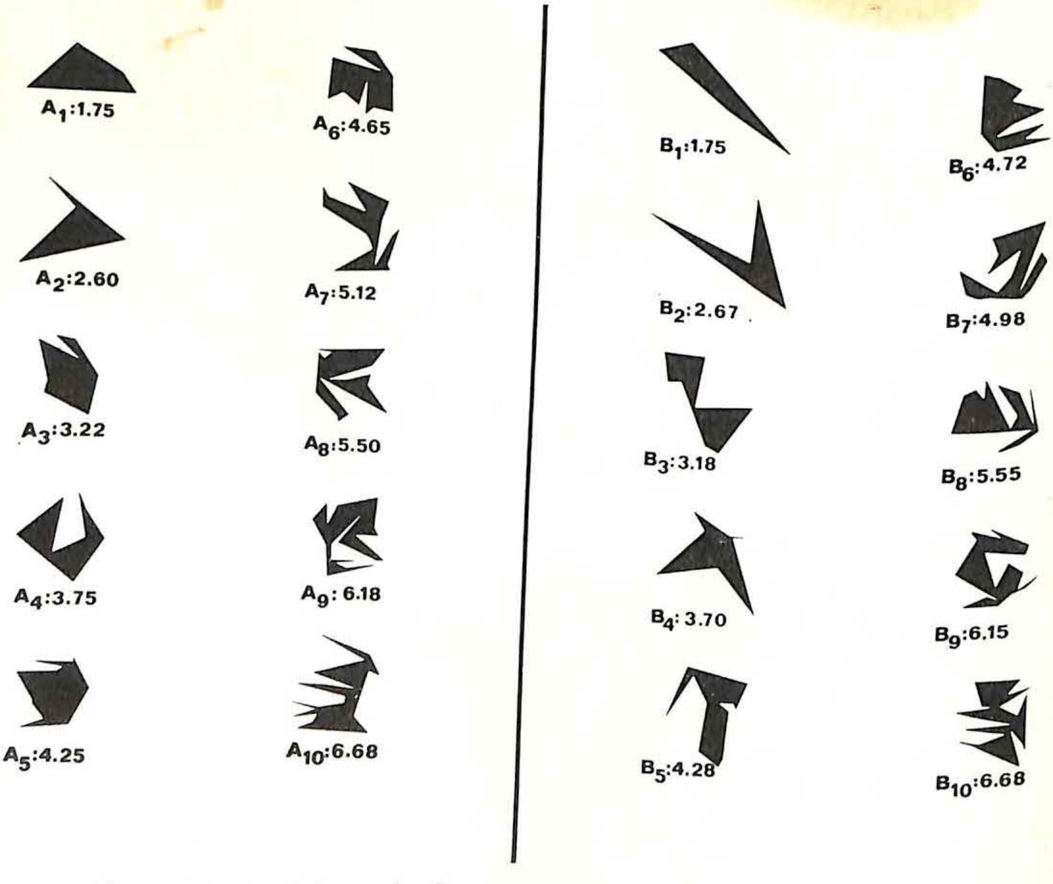


FIG. 8. Set A and B stimulus figures of Brown and Brumaghim with scaled complexity values. (Adapted from Brown & Brumaghim, 1968.)

TABLE 5

T_1, T_2 TOTAL FIGURE COMPLEXITY, AND SCALED COMPLEXITY VALUE FOR BROWN'S SET A AND SET B RANDOM POLYGONS

Stimulus figure	T_1 Lines	T_2 Angles	$\Sigma T_i/S_i$ Total figure complexity	Scaled value
A_1	12.7	10.0	22.7	1.75
A_2	11.6	31.4	43.0	2.60
A_3	29.3	41.2	70.5	3.22
A_4	26.6	45.9	72.5	3.75
A_5	66.4	73.3	139.7	4.25
A_6	46.7	69.8	116.5	4.65
A_7	96.6	97.1	193.7	5.12
A_8	89.9	95.3	185.2	5.50
A_9	93.4	106.5	199.9	6.18
A_{10}	96.4	120.8	217.2	6.68
B_1	11.9	23.2	35.1	1.75
B_2	9.9	35.0	44.9	2.67
B_3	35.1	32.6	67.7	3.18
B_4	29.4	61.8	91.2	3.70
B_5	65.2	125.0	125.0	4.28
B_6	49.4	74.3	123.7	4.72
B_7	75.3	84.5	159.8	4.98
B_8	90.6	92.5	183.1	5.55
B_9	96.9	105.7	202.6	6.15
B_{10}	90.6	132.3	222.9	6.68

generality of the two stimulus sets is hard to evaluate. This is not the case, however, for randomly generated shapes or polygons. Of direct relevance is the work of Brown and his colleagues. Specifically, Brown and Brumaghim (1968) have developed two sets of random polygons scaled for complexity. These two sets, A and B, are shown with their scale values in Figure 8.[6] Each set of 10 was selected from the results of scaling experiments using a large population of randomly generated polygons. The figures have 4, 8, 12, 16, or 20 sides and were selected to be approximately equidistant on a unidimensional visual complexity continuum. The present model should be capable of predicting these scaled complexity values. Table 5 presents the T_1, T_2, and $\sum T_i/S_i$ values predicted by the model and each polygon's associated scale value. Figure 9 shows the relation between the model's predictions and the scale values. In the case of Set A, the correlation between predicted total complexity and scale complexity value is .96; for Set B the correlation is .99. (Log_{10} of $\sum T/S$ was not used since the linear scale did not have the ceiling problem of the mean ranks found in Experiments I and II.)

Discussion

Although these correlations are quite high, the discrepancies do point to a weakness in the model. In Set A for No. 3 and 4, 5, and 6, and Set B for No. 5 and 6, the predictions are slightly off. In each of these cases, the two figures have the same number of sides and the model seems to miss because it underestimates the complexity of the figure with the more jagged or irregular outline. In retrospect we observed what appears to be the same problem with No. 19, 20, and 28 of Figure 4 in Experiment I and with No. 10, 15, and 17 of Figure 6 in Experiment II; that is, these figures also have very jagged outlines and have predicted complexity values that are somewhat too low. Perhaps elements based on the differences between adjacent lines and angles could be identified and sampled by the model. Something of this kind would have to be added in order to correct this problem.

The present model has several important advantages in comparison with other approaches to the prediction of figure complexity. It can be applied to any two-dimensional figure whether symmetric, partially symmetric, or asymmetric, and it predicts, without having run an earlier analysis of responses to a similar set of figures. Furthermore, as a model of the perceptual process, it is an integrated statement, however oversimplified, about how figure perception takes place, and therefore it makes predictions about other perceptual processes. For example, as described below, the model can be applied to various issues involving the perception of three-dimensional figures.

THE MODEL AND SUPPORTING DATA: THREE-DIMENSIONAL FIGURES

The importance of the third dimension for a model of figure or pattern perception

---

[6] D. R. Brown graciously provided examples of his original figures.

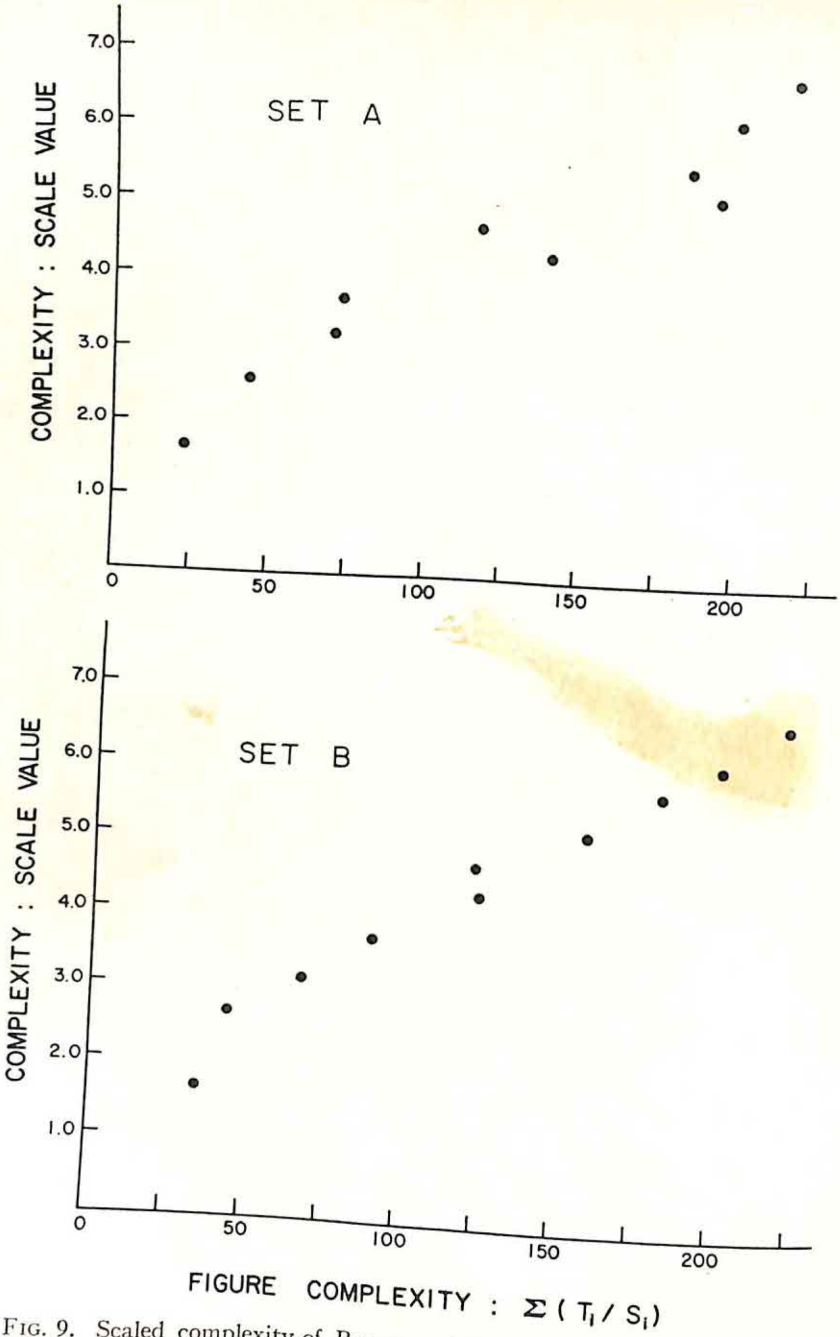


FIG. 9. Scaled complexity of Brown and Brumaghim's Set A and Set B figures as a function of predicted figure complexity. (Adapted from Brown & Brumaghim, 1968.)

is obvious from the three-dimensional character of most of everyday experience and from the problem posed by two-dimensional drawings which are perceived or recognized as three-dimensional objects.

Probably, the best known theoretical approach to the prediction of the three-dimensionality of a figure is that of Hochberg and McAlister (1953), Hochberg and Brooks (1960), and Attneave (1954). The basic argument is:

> the more information required to specify each alternative shape which may be reported in response to a given stimulus-figure, the less likely it is that

> that shape will be seen or reported relative to its alternates [Hochberg & Brooks, 1960, p. 339].

In short, the simplest figure out of a hypothetical array of possible figures will be the one perceived.

Hochberg and Brooks (1960) do not develop a rationale for identifying this hypothetical array of perceivable alternative figures. However, they do propose and present much supporting evidence for the following more specific hypothesis:

> the relative apparent tridimensionality of each member of a family of reversible perspective representations of a given three-dimensional object will be a simple function of the geometrical complexity of the two-dimensional figure [p. 340].

This interpretation follows from the basic argument, since as the two-dimensional complexity of a figure increases, it can be assumed that the alternative three-dimensional interpretations are relatively less complex. Examples of reversible perspective drawings of families of three-dimensional figures are shown in Figure 10. To test the hypothesis, Hochberg and Brooks developed three measures to predict perceived tridimensionality. The one measure of complexity which they found as the most suitable they denoted as Equation 3 and it equals: $[\sum\{T_2 + T_{12} + 2\,(T_4)\}]/N$. In this equation T_2 equals the total number of interior angles; T_{12} equals the total number of different angles; T_4 equals the total number of continuous line segments. N_T equals the number of the preceding tests which actually distinguish between the figures of the family being considered. (Note: each score of the above tests was also standardized on a 0–10 scale.) These properties of the equation and their particular weightings were based on a factor analysis of figure properties related to judged three-dimensionality. This measure has the advantage of easy computa-

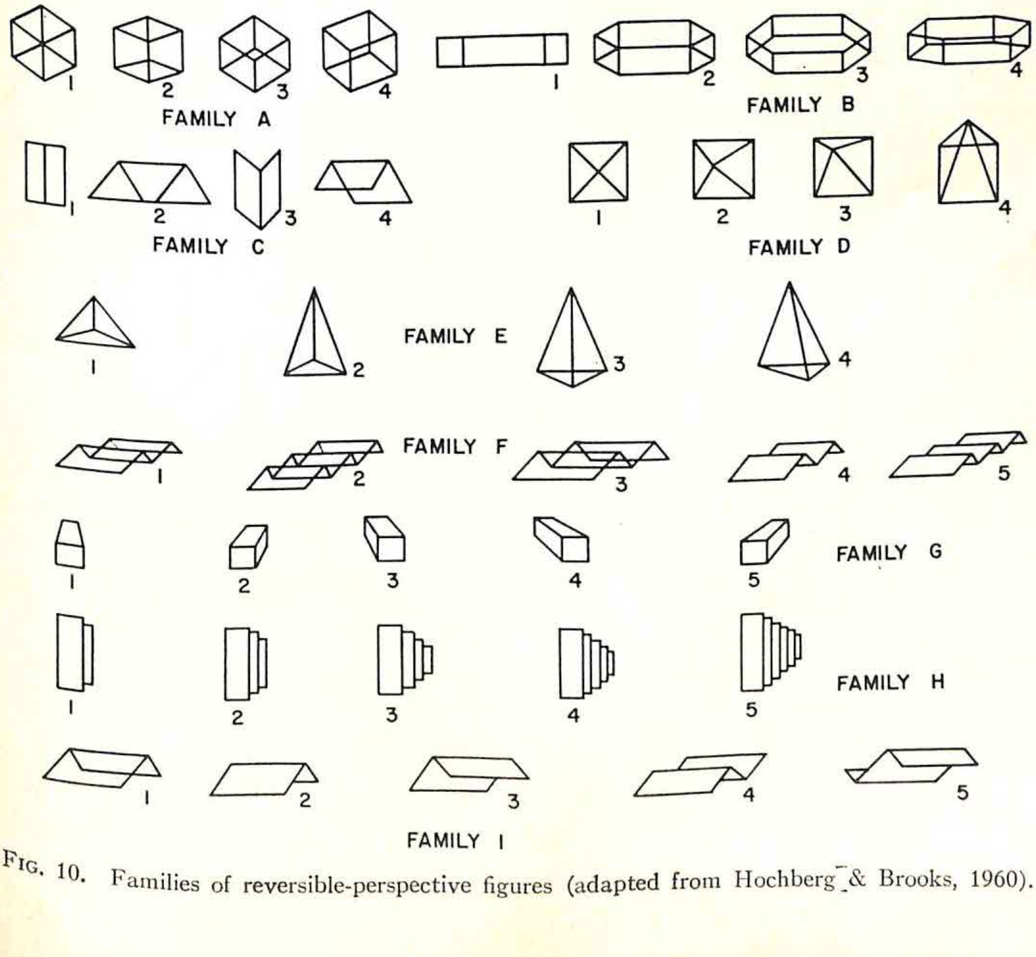


FIG. 10. Families of reversible-perspective figures (adapted from Hochberg & Brooks, 1960).

TABLE 6

TWO-DIMENSIONAL COMPLEXITY $\Sigma(T/S)$, HOCHBERG AND BROOKS MEASURE OF PERCEIVED TRIDIMENSIONALITY, AND JUDGED TRIDIMENSIONALITY FOR THE STIMULUS FIGURES SHOWN IN FIGURE 10

Family	(1) Equation No. 3 (Hochberg & Brooks, 1960)	(2) Complexity $\Sigma(T/S)$	(3) Judged three-dimensionality (Hochberg & Brooks, 1960)
A			
1	0.0	38.0	0.0
2	4.7	49.9	5.3
3	9.0	70.2	7.5
4	10.0	101.1	10.0
	$r_{13} = .98$	$r_{23} = .92$	
B			
1	0.0	28.1	0.0
2	4.8	74.2	3.7
3	4.4	108.3	10.0
4	10.0	168.3	10.0
	$r_{13} = .79$	$r_{23} = .91$	
C			
1	0.0	17.3	0.0
2	4.7	29.3	3.9
3	6.0	23.9	8.0
4	10.0	60.4	10.0
	$r_{13} = .98$	$r_{23} = .76$	
D			
1	0.0	25.1	0.0
2	5.0	72.0	7.6
3	7.5	88.5	9.1
4	8.2	87.4	10.0
	$r_{13} = .98$	$r_{23} = .99$	
E			
1	0.0	27.1	0.0
2	1.5	31.0	3.4
3	3.2	48.3	6.9
4	5.0	121.3	10.0
	$r_{13} = .99$	$r_{23} = .87$	
F			
1	1.7	174.9	0.0
2	7.5	273.0	3.5
3	4.3	284.7	3.9
4	2.5	97.0	6.9
5	2.5	181.2	10.0
	$r_{13} = -.13$	$r_{23} = -.31$	
G			
1	0.0	17.9	0.0
2	10.0	49.3	7.3
3	10.0	51.8	8.9
4	10.0	50.7	9.1
5	10.0	51.0	10.0
	$r_{13} = .97$	$r_{23} = .98$	
H			
1	2.5	25.8	0.0
2	3.1	50.4	4.3
3	4.4	71.9	6.6
4	6.0	102.4	8.4
5	7.5	137.7	10.0
	$r_{13} = .93$	$r_{23} = .95$	
I			
1	4.9	80.4	0.0
2	1.6	40.9	2.2
3	5.2	53.5	2.8
4	7.5	89.8	10.0
5	8.3	88.7	10.0
	$r_{13} = .79$	$r_{23} = .62$	

tion; however, it is not a generally applicable measure of two-dimensional complexity since it requires a figure to be a member of a family; nor is it developed with any special connection to an interpretation or theory of the process of perceptual organization.

The present model's measure of two-dimensional complexity can be used to predict perceived three-dimensionality. We will use the data from Hochberg and Brooks for this purpose and to explore the general problem of predicting perceived three-dimensionality. The stimuli in Figure 8 were shown by Hochberg and Brooks to college students who rated each member of the family of related figures on its degree of three-dimensionality.

> The *S*s were instructed to place the figure which seemed most strongly "solid" or "three-dimensional" at 10 on the scale, the "least solid" or "most flat" at 0, and then indicate where the other figures belonged between these two extremes. The means were standardized to a 0–10 scale within each family [Hochberg & Brooks, 1960, p. 341].

The obtained three-dimensionality values shown here in Table 6 were taken from Table 1, Test 1 of Hochberg and Brooks.

The results in Table 6 show that both complexity measures generally do well at predicting perceived three-dimensionality. Nevertheless, there are obvious failures, for example, Family F and I, and these suggest extending the model to incorporate the third dimension.

Before discussing the addition of the third dimension to the model, a few words about the stimuli in Figure 10. Family A represents the well-known Necker or Kopfermann cubes. These and most of the remaining figures are reversible figures, that is, have two structurally identical but perceptually different locations in space. All of the figures within a family are ordered in terms of increasing judged tridimensionality. (Figure 10 was taken from Plate 1 of Hochberg and Brooks. There are some figures which were incorrectly drawn in the original plate, these have been changed to coincide with the actual stimuli.)[7] A model of figure perception should be able to easily predict reversibility. In addition, the various alternative perceptions of a pattern should have some special existence within the model. For example, No. 1 of Family A (Figure 10) can be seen as a three-dimensional but almost flat figure with the alternating sets of three triangles appearing above and then below the surface. It is possible to interpret it as the top view of a six-sided pyramidal solid. No. 2 of Family C also has at least two interpretations in three-dimensionality. It can be seen as a side view of a figure consistent with the other members of the family—or the center trapezoid shape can be seen as a front plane and the two triangles as part of a larger back plane also of a trapezoidal shape.

ADDING THE THIRD DIMENSION: THE EASY WAY

The arguments behind the previous definition of elements and their properties is easily extended to define volumes as nonoverlapping and completely enclosed spaces formed by the intersection of planes, which are sampled in proportion to their size. With the extra dimension, there are three possible axes of symmetry and S takes on a value of 4.00 for most of the stimuli in Figure 10.

Furthermore, the orientation of lines in three dimensions, necessary for the specification of correspondence in evaluating sampling redundancy, is specified by the orientation of the line relative to the plane defined by the horizontal and vertical axes and by the orientation of the line relative to the plane defined by the vertical axis and the axis of the third dimension. Thus, to specify orientation of a line in three dimensions, two angles are necessary. One can be thought of as the orientation of the projection of the line onto the x–y plane, defined in the two-dimensional sense, and the other as the orientation relative to the y–z plane. Correspondence in three dimensions requires identity on both the x–y and y–z orientation.

Given the figures' element properties, the model can be used easily to evaluate a three-dimensional complexity. How to get these properties is a major problem, however. For the present, we will use the weak approach of using the information from all figures in a family, and we will assume the figures are all of the same three-dimensional object. For example, No. 1 of Family I is interpreted as two rectangular planes approximately 2×1 in size and meeting at a 90° angle. Ambiguity about angle sizes is resolved by the additional simplifying assumption that as many angles as possible are assumed to be equal, for example, for Family I, angles are only 90°. The original hypothesis can be used to argue that the *difference* between a figure's two- and three-dimensional complexity should be a better predictor of perceived dimensionality than two-dimensional complexity by itself. It is the relative simplicity of the two-dimensional and three-dimensional representation that counts. For most of the stimulus families in Figure 10, this interpretation will not change matters, since each figure in the Family has the same three-dimensional complexity, and subtracting a constant will have no effect. This is not, however, the case for Family I, since No. 1 in Family I is reversible unlike the rest of its family; also in Family F, only No. 1, 2, and 3 are reversible.

---

[7] We are indebted to Julian Hochberg and Virginia Brooks for providing copies of their original stimulus figures.

An important problem is to estimate how much more complex a reversible three-dimensional figure is compared to the "ordinary" nonreversible variety. The model is based on the assumption that a reversible three-dimensional figure is weighted, that is, multiplied, by the number of three-dimensional angles in the figure which have two three-dimensional interpretations. By a three-dimensional angle is meant the angles created at the intersection of two or more planes. For example, a Necker cube has eight corners, resulting from the intersection of three planes, and each corner has three three-dimensional angles having two three-dimensional interpretations; hence its reversibility weighting factor is 8 × 3 or 24. The reversibility weighting factors for Families F and I are shown in Table 7, column 7. For No. 1 of Family I (Figure 10), the intersection of two planes has two corners with three angles at each; and the number of reversible three-dimensional angles (reversibility weighting factor) is 12, for No. 3 of Family I, the reversibility weighting factor is 18, etc.

Using this rationale for reversible figures, the model's new predictions of perceived three-dimensionality, shown in Table 7, are appreciably better. The correlation between predicted and judged three-dimensionality for Family F changes from −.31 to .53; for Family I, from .62 to .95.

ADDING THE THIRD DIMENSION: THE HARD (AND UNSOLVED) WAY

The central issue is how to represent the three-dimensional figure given the two-dimensional drawing. In the preceding treatment we used information from the family of related figures but, of course, much of the time, this information is not present. Also, to assume all angles are equal if they can be equal is a questionable assumption. How does one know in a rigorous fashion if it is possible for the angles of a figure to be equal and consistent with the two-dimensional information? Furthermore, many figures which could have equal

TABLE 7

THE TWO-DIMENSIONAL (2-D) COMPLEXITY MINUS WEIGHTED THREE-DIMENSIONAL (3-D) COMPLEXITY FOR THE FIGURES OF FAMILY F AND I

Family	2-D complexity (from Table 6) (1)	3-D elements			S	3-D complexity $\Sigma T_i/S_i$ (6)	Reversibility weighting factor (7)	3-D $\times$RFW[a] (8)	2-D minus weighted 3-D complexity (9)	Judged 3-D (from Table 6) (10)	New r 9–10
		T_1 (2)	T_2 (3)	T_3 (4)	(5)						
F											
1	174.9	49.1	77.3	8.4	4.00	31.2	18	561.6	− 386.7	0.0	$r = .53$ 9–10
2	273.0	76.1	127.0	15.0	4.00	54.5	30	1635.0	−1362.0	3.5	
3	284.7	49.1	77.3	8.4	4.00	31.2	18	561.6	− 276.9	3.9	
4	97.0	49.1	77.3	8.4	4.00	31.2	1	31.2	65.8	6.9	
5	181.2	76.1	127.0	15.0	4.00	54.5	1	54.5	126.7	10.0	
I											
1	80.4	21.6	29.4	3.0	4.00	13.5	6	81.0	− 0.6	0.0	$r = .95$ 9–10
2	40.9	21.6	29.4	3.0	4.00	13.5	1	13.5	27.4	2.2	
3	53.5	21.6	29.4	3.0	4.00	13.5	1	13.5	40.0	2.8	
4	89.8	35.0	53.4	5.6	4.00	23.5	1	23.5	66.3	10.0	
5	88.7	35.0	53.4	5.6	4.00	23.5	1	23.5	65.2	10.0	

[a] RWF = Reversibility weighting factor.

angles obviously do not look that way. For example, slightly "squashed" Necker cubes are quite possible. The reader may have already noticed the slightly distorted properties of the Necker cubes, and of other stimuli in Figure 10. In short, there are an infinite number of rational or logically possible three-dimensional interpretations of a given two-dimensional drawing. The problem is to find those underlying assumptions of the perceptual process which reduce the possible interpretations to one or to very few at most. In an intuitive sense, those three-dimensional interpretations which are perceived seem to be especially simple. The idea of reducing the possible perceived figures to the simplest requires that "simplest" be thoroughly described or defined. As a first approximation, the simplest figure appears to be that figure with the fewest number of angles and with the least variability in angle size. An important move toward conceptualizing such a minimum principle has recently been published by Attneave and Frost (1969). Another clue to solving the three-dimensional representation problem is contained in the recent work of Hochberg (1968) which very clearly identifies the importance of angles created by line intersection as being fundamental units in the perceptual construction of the three-dimensional figure.

In summary, the model's two dimensional complexity values do moderately well at predicting rated three-dimensionality; the difference between a figure's two- and three-dimensional complexity predicts somewhat better. However, the model suffers seriously from the absence of a general rationale for interpreting the three-dimensional representation of a figure from its two-dimensional drawings.

REFERENCES

ATTNEAVE, F. Some informational aspects of visual perception. *Psychological Review*, 1954, **61**, 183–193.

ATTNEAVE, F., & FROST, R. The determination of perceived tridimensional orientation by minimum criteria. *Perception and Psychophysics*, 1969, **6**, 391–396.

BITTERMAN, M. E., KRAUSKOPF, J., & HOCHBERG, J. Threshold for visual form: A diffusion model. *American Journal of Psychology*, 1954, **67**, 205–219.

BORG, G. Studies in visual gestalt strength. Unpublished manuscript, Umeå University, Umeå, Sweden, 1964.

BROWN, D. R., & BRUMAGHIM, S. H. Perceptual equivalence, pattern perception and multidimensional methods. *Perception and Psychophysics*, 1968, **4**, 253–256.

CAMPBELL, F. W., KULIKOWSKI, J. J., & LEVINSON, J. The effect of orientation on the visual resolution of gratings. *Journal of Physiology*, 1966, **187**, 427–436.

DODWELL, P. C. Coding and learning in shape discrimination. *Psychological Review*, 1961, **68**, 373–382.

GANZ, L. Analysis of pattern perception through visual deprivation. In P. C. Vitz (Chm.), Pattern perception. Symposium presented at the meeting of the American Association for the Advancement of Science, Boston, December 1969.

HANES, R. M. Some effects of shape on apparent brightness. *Journal of Experimental Psychology*, 1950, **40**, 650–659.

HEBB, D. O. *The organization of behavior.* New York: Wiley, 1949.

HIGGINS, G. C., & STULTZ, K. Visual acuity as measured with various orientations of a parallel-line test object. *Journal of the Optical Society of America*, 1948, **38**, 756–758.

HOCHBERG, J. In the mind's eye. In R. N. Haber (Ed.), *Contemporary theory and research in visual perception.* New York: Holt, Rinehart & Winston, 1968.

HOCHBERG, J., & BROOKS, V. The psychophysics of forms: Reversible perspective drawings of spatial objects. *American Journal of Psychology*, 1960, **73**, 337–354.

HOCHBERG, J., & MCALISTER, E. A quantitative approach to figural "goodness." *Journal of Experimental Psychology*, 1953, **46**, 361–364.

HUBEL, D. H., & Wiesel, T. N. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Journal of Physiology*, 1962, **160**, 106–154.

HUBEL, D. H., & WIESEL, T. N. Receptive fields and functional architecture in two nonstriate visual areas (18 and 19) of the cat. *Journal of Neurophysiology*, 1965, **28**, 229–289.

KAUFMAN, L., & RICHARDS, W. Spontaneous fixation tendencies for visual forms. *Perception and Psychophysics*, 1969, **5**, 85–88.

KEMENY, J. G., & SNELL, J. L. *Finite Markov chains.* Princeton, N. J.: Van Nostrand, 1960.

MCFARLAND, J. H. Sequential part presentation: A method of studying visual form perception. *British Journal of Psychology*, 1965, **56**, 439–446.

MCFARLAND, J. H. "Parts" of perceived visual forms: New evidence. *Perception and Psychophysics*, 1968, **3**, 118–120.

PRITCHARD, R. M., HERON, W., & HEBB, D. O. Visual perception approached by the method of stabilized images. *Canadian Journal of Psychology*, 1960, **14**, 67–77.

RICHARDS, W., & KAUFMAN, L. "Center of gravity" tendencies for fixations and flow patterns. *Perception and Psychophysics*, 1969, **5**, 81–84.

RIGGS, L. A., & TULUNAY, S. U. Visual effects of varying the extent of compensation of eye movements. *Journal of the Optical Society of America*, 1959, **9**, 741–745.

USPENSKY, J. V. *Introduction to mathematical probability*. New York: McGraw-Hill, 1937.

VANDERPLAS, J. M., & GARVIN, E. A. The association value of random shapes. *Journal of Experimental Psychology*, 1959, **57**, 147–154.

VITZ, P. C., & TODD, T. C. A coded element model of the perceptual processing of sequential stimuli. *Psychological Review*, 1969, **76**, 433–449.

(Received April 22, 1970; revision received December 23, 1970)

Psychological Review
1971, Vol. 78, No. 3, 229–248

THE DYNAMICS OF STRUCTURED PERSONALITY TESTS: 1971 [1]

DOUGLAS N. JACKSON [2]

University of Western Ontario

Issues raised by Paul Meehl's classic manifesto on the dynamics of structured personality inventories are reviewed in the light of a quarter of a century of experience. A series of principles relating to personality scale development is proposed, which highlights the role of psychological theory and item content in formulating a multifaceted item pool, and the importance of suppressing response biases. Because of the competing and often interacting relation of trait and method variance, multivariate procedures for scale construction and evaluation of discriminant validity are considered essential. It is concluded that considerably more is now known about personality and scale construction than 25 years ago. Personality assessment specialists need not, and indeed as scientists should not, abandon their unique human capacity to judge and evaluate item content.

It is a quarter of a century since Paul Meehl (1945) published his now classic manifesto on the dynamics of structured personality inventories. Meehl's article ushered in a very important new era in personality assessment, with perhaps just a bit of justifiable confidence and even missionary zeal. Those doing research in the area of personality assessment owe a tremendous debt of gratitude to Paul Meehl and the many people who followed in this tradition. Tremendous enthusiasm was generated for the Minnesota Multiphasic Personality Inventory (MMPI), stimulating confidence in personality assessment and research. Equally important were the many research questions this procedure raised. Questions relevant to the determinants of responses to personality items, actuarial versus clinical prediction, the degree to which the structure of personality is revealed in responses to personality inventories, and the sources of bias that may enter into both the individual scores and the interpretation of scores are just a few of the many fertile areas of research stemming at least in part from Meehl's original article.

The purpose of this paper is to survey certain theoretical developments in personality assessment since the publication of Meehl's article. Controversial issues will be raised regarding much of what Meehl and some of his followers have proposed. But it is hoped that out of discussion of these issues there may ensue a richer and deeper understanding, particularly if the search for knowledge is conducted in the spirit that Meehl so strongly advocated, that is, a willingness to be guided by empirical data and to correct one's conceptions and misconceptions on the basis of empirical findings.

In the past quarter of a century, there has been reason to be justly proud of some very real achievements in personality measurement. Some of these contributions, like the work of Meehl and Hathaway (1946) represent important milestones. In a manner similar to Sir John Suckling's "Constant Lover," personality assessment has been like an adolescent seeking romantic infatuations with a variety of novel developments. But many contributions, like early loves, have a lasting impact. Notable among these contributions turned milestones have been the introduction into assessment of questions of substance by Cronbach and Meehl (1955)[3] and by

---

[1] Portions of this paper were presented as an invited research address to the meeting of the Western Psychological Association, Vancouver, B. C., June 1969. This research was supported by Research Grant No. 151 from the Ontario Mental Health Foundation. Grateful acknowledgement is extended to Kenneth A. Carlson and John A. Neill for a careful reading of the manuscript.

[2] Requests for reprints should be sent to Douglas N. Jackson, Department of Psychology, University of Western Ontario, London 72, Canada.

[3] It is noteworthy that Meehl's recognition of the importance of theory to personality assessment as contained in the important construct validity article

Loevinger (1957); the distinction between convergent and discriminant validity and method and trait variance, explicated by Campbell and Fiske (1959); and the identification by Cronbach (1946, 1950) and others of response sets and styles followed by a vigorous and controversial research literature regarding the identification, role, and correlates of response style. Another milestone is not indicated by a single article or single contribution, but by the cumulative weight of a great many investigations. Reference is made here to the impact of test and measurement theory, of factor analysis, and of the whole class of multivariate techniques upon present thinking in personality. The field of psychometrics has been very much a growing, cumulative field of science. It has provided powerful techniques, which, when coupled with the modern computer, afford a degree of power and precision that simply did not exist in 1945. The opportunities for applying these techniques for further developments in the psychometric or information-processing tradition should not be overlooked. In fact, they should be capitalized on whenever possible.

Despite their already acknowledged contributions, none of these milestones has provided any sort of touchstone to the measurement of personality, primarily because each has not been pushed to its logical extension and because investigators have been too little aware of the full implications and the full integration of these various approaches. Stated positively, much more knowledge about the nature of personality assessment is available than the fruits of some fragmented developments might imply, and an integration of developments will reveal possibilities for some truly significant forward thrusts. This paper focuses on a set of principles derived in part from these milestones upon which personality assessment theory and method may proceed.

THE CASE FOR EMPIRICAL ITEM SELECTION

Since Paul Meehl's manifesto regarding the construction of personality tests is taken as a point of departure, it might be appropriate to review Meehl's position. Meehl defined structured personality tests as those in which the test material consisted of conventional culturally crystalized questions to which the subject must respond in one of a few fixed ways. Meehl also suggested that there were essentially two approaches to assessment, one which he called the traditional approach and one which he characterized at that time as a modern approach because it provided, in his words, more sophisticated views of the assessment situation. The first approach assumed that a person's responses to a structured personality test involved a kind of sample of behavior, a "second best source of information when direct observation was unavailable." The essence of this approach was that the investigator on intuitive grounds defined the content and sought to identify verbal behavior in a way which would permit him to make some kind of inference about the traits. Meehl rejected this approach as inopportune for a number of reasons. It is possible to agree heartily with Meehl's assessment of the quality of many attempts at personality scale construction prior to 1945, while disagreeing with his attribution of the cause. In this paper an attempt is made to show that a sampling conception of an attribute, when properly applied, is the preferred approach in the great majority of instances.

The second approach, according to Meehl, was that a person's response to a questionnaire item was in itself a basically interesting datum of behavior whose correlates must be uncovered empirically. It was not for the investigator to say what this behavior meant, but rather this was something that should and must be deduced from the empirical correlates of the responses to the items. He used as his example the experience with the Strong Vocational Interest Blank (SVIB) and the MMPI. Both of these devices, at that

---

has substantially influenced the present formulation. Controversial issues raised in the present paper should therefore be construed as referring to certain interpretations of the 1945 Meehl article and not to Meehl's present thinking on these issues.

time, were relatively new.[4] Both showed great promise. They were themselves milestones in that they were among the first multiscore structured personality tests. Both used the method of empirical item selection in their construction, and both purported to represent comprehensive, if not universal, characterizations of the domains which they sought to approach; namely, vocational interests and psychopathology. Both were developed in an intellectual climate of the radical empiricist reaction against earlier philosophical traditions in psychology. Although this reaction was, I believe, both necessary and healthy, it did involve an unverbalized belief that rigorous science is equated with the rejection of theory.

Perhaps as a consequence of such beliefs, it was almost 30 years since the development of these two devices before substantial interest developed in systematically studying item content as it might be linked to scale scores, correlates, or criterion groups. While there were some important exceptions, among many adherents it was almost considered unscientific or mystical to investigate categories of content. Certainly, individuals who followed in this tradition received support from Meehl's (1945) article, if not from his more recent contributions (Cronbach & Meehl, 1955). Meehl suggested that the use of empirical item selection did not require every subject to view the item content in the same way. The hypochondriac, for example, might well interpret the word *frequently* in the item, "I have headaches frequently," as something that is more regular than the nonhypochondriac. Thus, according to Meehl, the nonhypochondriac might have headaches just as frequently, but the critical issue was that the hypochondriac interprets the item differently. This illustrates perhaps a most general point in the empiricist tradition. That is, the scoring of an empirically derived inventory such as the MMPI does not require accurate self-rating to be useful. In Meehl's own words,

> What a man says about himself may be a highly significant factor, even though we do not entertain with any confidence the hypothesis that what he says would agree with what complete knowledge of him would lead others to say of him [p. 299].

A corollary of this approach is that items completely lacking in any rationale, or items which might even appear to be opposite from what might be expected from theory, are permitted with this technique. Thus, Meehl suggested that there is no good rationale why an individual who might be characterized clinically as a psychopathic personality would respond true to the item, "My parents and family find more fault with me than they should"; surely as a *fact* such a characteristic would be difficult to validate. But because this item has been shown to discriminate empirically between such a criterion group and normals, it is considered useful in diagnosis. A number of other even more psychologically remote examples of items keyed on MMPI scales, apparently bearing no substantive relationship to the dimension of psychopathology putatively represented by the scale, were introduced. These items were considered to have had positive value, in that they provided a foundation for overcoming faking and distortion by providing content which even an expert in psychopathology had no good rationale for linking to the nature of the disturbance.

Since the approach advocated here departs significantly from the empirical tradition, there will be occasion to return to many of these points, as well as to many of the contributions alluded to by other research workers during this era. An attempt is made to demonstrate that what might appear to be sources of controversy and potential bases for argument may actually be integrated into a formulation of the dynamics of structured personality inventories. This formulation, while incomplete and partial, may serve to offer a comprehensive basis for moving forward with the valid, substantively meaningful assessment of personality.

---

[4] Because the first SVIB manual was published in 1928, it may not be entirely accurate to characterize it as "relatively new" in 1945. However, in the early years, recognition for Strong's achievement was slow in developing, but had taken firm hold by the 1940s.

IMPORTANCE OF THEORY IN SCALE CONSTRUCTION

The first general principle is that *personality measures will have broad import and substantial construct validity to the extent, and only to the extent, that they are derived from an explicitly formulated, theoretically based definition of a trait.* This principle is based on the broad assumption that every act and every criterion performance will usually reflect a wide range of underlying traits. But it will not reflect each trait with equal power or precision. Thus, because every behavior, verbal or otherwise, is a manifestation of some underlying set of characteristics, these characteristics may be assumed to be measurable by a sampling of behavior. But it will not do to sample just any behavior. Assessing that set of behaviors most relevant to the trait in question is the most efficient way of measuring the trait. Ordinarily, issues, questions, or situations directly bearing on the characteristics of interest will show the highest probability of reflecting them. Thus, for example, if we wish to measure a trait such as dominance, items bearing directly on characteristic modes of leading or directing other people and derivatives therefrom would be more likely to be reliable and valid indicators of such a trait than would more remote, substantively oblique manifestations of dominance. Certainly, one of the least efficient ways of proceeding to measure such a trait would be to compile randomly a heterogeneous, unselected group of items and then to derive empirically a handful of items differentiating criterion groups.

Of course, a substantive approach places a certain responsibility on the scale developer. His responsibility is to have some understanding of what it is he wishes to measure. He must not be content with the mere prediction of a criterion without knowledge of the psychological makeup of the characteristics involved. But it is doubtful whether very many psychologists are solely concerned with the prediction of a criterion. Probably few of us use psychological tests in the classical manner of the applied psychologist, who might have been concerned with such questions as predicting which salesman will sell the most insurance. Few are even interested in predicting how a psychiatrist will diagnose a patient. Rather, most of us are primarily interested in imputing latent characteristics to respondents on the basis of their test scores. Even the MMPI and the SVIB, both originally empirically derived, are used typically to form some characterizations of underlying traits. Indeed, the recent revision of the SVIB (Campbell, 1969), in a radical departure from previous practice, now recommends that a new set of rationally constructed scales form the major basis for interpretation.

Cronbach and Meehl (1955) have suggested that empirically derived scales might serve to enrich understanding by a bootstrapping technique, much as in the manner of Alfred Binet, who, when he started, purportedly knew little more about intelligence than was contained in teachers' criterion ratings of bright and dull pupils. But such a procedure is justified only under circumstances of complete or almost totally complete ignorance. Ordinarily, psychological science will be served better if an investigator is willing to take a stand on the nature of a construct, to define it, to measure it, and to test its implications in a manner that allows him to confirm or disconfirm his hypotheses. There was a point when we were psychologically naive about the areas of psychopathology and of vocational interest. It is perhaps in one sense fortunate that Edward K. Strong and Hathaway and McKinley proceeded in the way they did at a time when the issue of whether or not vocational interests and psychopathology could be measured at all was a matter of some dispute. But the ignorance about personality to which Meehl alluded a quarter of a century ago hardly seems a suitable defense at the present time. Research studies in psychology account for one of the largest sources of published scientific literature, and a substantial portion of this research concerns areas related to personality. There is even considerable evidence that a nonpsychologically trained individual can make quite refined distinctions regarding personality, if the questions are properly put to him.

To support the assertion that there is an overwhelming substantive basis for personality item responding, data from two general sources will be drawn upon, the first an analysis of content dimensions in the Personality Research Form (PRF) and in the SVIB, and the second derived from results of studies of judgmental processes in the perception of personality and personality scale content.

During the course of the construction of the PRF (Jackson, 1967c), one prerequisite to preparing item pools was a very careful review of the definitions of variables of personality given by Murray (1938), which in many cases also involved an elaboration and reformulation of these definitions. These definitions were then translated into a large sample of items, designed to serve as a comprehensive representation of the characteristic, removing sources of bias insofar as was possible, and writing items with critical regard to the conceptual overlap between scales. On several occasions, we had recourse to both formal psychological theory and to more informal substantive elaborations of the trait. This is precisely the kind of rational process about which empiricists have been so skeptical, and about which Meehl cautioned a generation of test constructors. We did not, of course, stop at merely a substantive appraisal of the dimension, but sought to evaluate this procedure in a number of ways. The first important evaluation stemmed directly from the item-analysis procedures. Listed on Table 1 are 7 of the 20 substantive PRF scales, together with the initial number of items written for each scale. These items had already been subjected to thorough editorial review and revision, so that a good many items which were initially written for the scale were not even included in the item analysis for failing to pass the first substantive hurdle. These items were administered to 309 male and female university students in two sessions separated by one week, a procedure necessitated by a concern for the endurance it would have required to answer almost a thousand personality items in one sitting.

TABLE 1

COMPARISON OF IRRELEVANT VERSUS PREDICTED ITEM-TOTAL SCALE BISERIAL CORRELATIONS IN THE PERSONALITY RESEARCH FORM ITEM POOL

Scale	No. of items	K-R 20	$\bar{r}_{ig}$	Percentage $\bar{r}_{ig} > \lvert r_{is} \rvert$
Dominance	136	.93	.66	100.0
Autonomy	140	.89	.56	99.9
Understanding	144	.92	.55	100.0
Affiliation	130	.93	.70	99.9
Change	138	.86	.48	99.8
Nurturance	133	.93	.63	99.8
Impulsivity	134	.92	.62	100.0

Note.—$N = 309$.
$\bar{r}_{ig}$ = mean biserial correlation between an item and its own total scale for the best 40 items selected on the basis of the Differential Reliability Index (DRI); DRI = $\sqrt{r_{ig}^2 - r_{iD_y}^2}$, where r_{iD_y} is the biserial correlation between the item and a desirability scale; $\lvert r_{is} \rvert$ = the absolute value of the correlation between an item and an irrelevant scale.

First, an examination of the Kuder-Richardson Formula 20 reliabilities on Table 1 reveals that these are as high as those for some of the better achievement tests of comparable length. If any one should suggest that personality assessment has not yet achieved the level of fidelity attained in achievement testing, it might be appropriate to point out that comparable levels of reliability can be attained, and that, in addition, more is usually demanded of a personality test. Few achievement test specialists, for example, demand that their procedures demonstrate convergent and discriminant validity. The level of reliability attained encouraged further evaluation of item properties. Biserial correlations were computed between each item and its own total scale, together with correlations with each irrelevant scale and a specially devised desirability scale. For each scale, the best 40 items were chosen on the basis of a formula which seeks to maximize the content saturation of the item in relation to its saturation with a desirability scale, the Differential Reliability Index, developed in collaboration with Samuel Messick (cf. Neill & Jackson, 1970). Listed in Table 1 are the mean biserial correlations ($\bar{r}_{ig}$) of the best 40 items for each scale with its own total scale.

It is of interest to inquire to what extent irrelevant items, that is, items written for a different scale, correlated more highly with a particular scale than did the mean of the 40 items finally chosen for inclusion in Forms A and B of the PRF. Hypotheses about traits implicit in each item can thus be evaluated, providing an index of the degree of confusion between personality constructs. The number of times that this did not occur in relation to the total number of opportunities for the intrusion of an irrelevant item can be expressed as a proportion as indicated in the final column in Table 1. For the Dominance scale, there were no occasions when 1 of the 819 irrelevant items exceeded the mean content saturation of the best 40 Dominance items. This perfect batting average occurred for the Understanding and Impulsivity scales as well. For the other scales, the percentage of hits was almost as high; in no case was it lower than 99.8%. For the entire pool of items, there were approximately 5,700 opportunities for irrelevant intrusions of the type described, but these occurred only a total of five times, and only as often as twice in any one scale. This is strong evidence, I believe, for the proposition that a group of item writers can have more than trivial knowledge about the nature of personality constructs, and can proceed from a position other than from one of sheer ignorance. It should be emphasized in this comparison that these items were not simple restatements of the same theme, but covered the gamut of situations and modes of response appropriate for sampling broadly a personality construct.

It would not be accurate, however, to imply that we were never wrong in our miniature hypotheses contained in each item about a particular personality construct. On a number of occasions, items did not show the high levels of content saturation that were demanded, and therefore sometimes manifested an inappropriate pattern of correlations with irrelevant scales. On other occasions, an item correlated too highly with the desirability scale, or showed disproportionate endorsement frequencies. But in the aggregate, hypotheses were much more frequently confirmed than disconfirmed. The subjecting of a hypothesis to empirical test was a chastening as well as a rewarding experience, in that a great deal was learned about how complexities and ambiguities of wording could extract a toll. But it is infinitely more efficient to proceed by postulating explicitly the nature of that which was to be measured than by starting with an unselected item pool garnered from any convenient source. Furthermore, a substantive basis for item selection tends to lessen the chances of initially selected items failing when administered to a new sample of respondents. It is my hunch that the great majority of subtle items uncovered with the MMPI (Wiener, 1948), and considered a positive virtue by the proponents of empiricism, are present in MMPI scales due to errors in sampling items and subjects in the initial item-selection procedures of the MMPI. Most subtle items have been shown to correlate negatively with the rest of the items contained in a particular MMPI scale, raising the suspicion that they did not belong there in the first place.

The above assertions imply that a sample of item content most directly relevant to a particular trait will be the most efficient way of going about measuring it. Ordinarily, this assertion has a great deal of merit, but it must be amended in certain particulars. If an investigator should want to measure sadistic impulses, for example, it would be quite improbable that he would observe a person torturing animals, nor would any but a very small minority of people endorse item content regarding the satisfaction they derive from torturing animals. Of course, if such behavior were to be observed, an inference about sadistic impulses would be quite defensible. But ordinarily it would be necessary to incorporate subtlety and relatively more remote characterizations of such impulses. For example, items inquiring into a preference for the occupation of butcher or into the person's attitudes about the punishment of criminals might be appropriate. For most traits of personality and psychopathology, it is possible to identify behavior capable of being incorporated into item content in such a way as to avoid arousing such a degree of defensiveness as to render valid assessment unattainable.

The SVIB is considered by many as one of the most successful psychological tests and also the finest example of the use of empirical item selection with criterion groups. Perhaps because of the rationale employed by the original test author (cf. Campbell, 1968), there was a general reluctance to look at item content, and for many years the item content in the SVIB was not considered a fit matter for investigation. Recently, however, David Campbell and his colleagues at the University of Minnesota (Campbell, Borgen, Eastes, Johansson, & Peterson, 1968) proceeded to construct by a combination of rational and correlational means a set of content keys for the SVIB, covering such areas as music, art, and mathematics, and some two dozen others. It was possible for these investigators to employ the very extensive data bank of the responses of a diversity of occupational groups accumulated on the SVIB. A perusal of the results of a variety of occupational groups reveals that in every single instance, a criterion group substantively related to the particular content scale obtained the highest score on that scale, and that the other groups were ranked according to their relation to the activity. Thus, mathematicians expressed a liking for such activities as calculus and algebra more than did any other group. Musicians expressed a preference for musical activities, salesmen for selling, and artists for art. This should not be at all surprising, but perhaps it might be to someone overly committed to the empirical tradition. Rather, it is perhaps surprising that so little attention was focused on this aspect of vocational interest preference.

Returning now to personality assessment proper, a critic might wish to challenge the PRF item-analysis findings on the grounds that the item writers at Stanford University where the work was done were a distinctly talented group who possessed very much more than an average degree of psychological training and insight. This may have been true, but, at the same time, there is evidence to indicate that this in no way diminishes the important role played by substantive considerations in responding to a personality item. There are data indicating that subjects of no more than average sophistication in the psychology of personality have shown remarkable acumen. Much of this work derives from investigations of trait inference theories utilized by judges in studies of person perception.

In one such study (Jackson, 1970), a number of hypothetical people were characterized in terms of a narrative description. One hypothetical person was prepared to represent each pole of three personality dimensions, Autonomy, Impulsivity, and Dominance. For example, the description below represented a person at the positive pole of Autonomy.

> *Edward Archer* is a free-lance writer. At one time he worked for a newspaper, but quit because he felt restricted by the company's regulations. He now enjoys his work because he is his own boss. Edward spent the last year travelling alone through several European countries. He has currently taken up temporary residence in a large city's Bohemian district. Edward strongly advocated the view that today's young people are mere imitations of each other. He frequently tells people that the world needs more "individuals." His friends describe him as independent, headstrong, and freedom-loving.

Items from each of the three PRF scales were also selected to represent each pole of these dimensions. Judges, 59 officer candidates from the Canadian Forces, were given the task of judging the probability that the particular hypothetical person described would endorse each of these statements, and, in addition, of judging the joint probabilities that pairs of statements would be endorsed in a similar direction by a person. Judges were divided randomly into two groups, and completely distinct multidimensional scaling analyses were undertaken for these two groups. Three dimensions were extracted and rotated analytically to a Varimax criterion. In each case, the appropriate hypothetical person was salient on the appropriate pole of each of the three hypothesized dimensions, and all of the appropriate PRF items were projected on the appropriate dimensions with no exceptions and no nonpredicted items appearing where they should not. On the separate analyses, the randomly divided groups showed convergence for the three dimensions as indicated by product-moment correlations for the three respective dimensions of .99, .99, and .98. Clearly, when the questions are properly put, judges can reliably evaluate the underlying content dimensions contained in a set of personality items.

Results from the above study are consistent with findings recently published by Lay and Jackson (1969) of a very high degree of agreement in multidimensional scaling analyses of trait inferential judgments of male and female judges. Furthermore, it was found that the implicit "mapping" of traits, as, for example, in a finding of a negative inferential relation between an Order statement and an Impulsivity statement, were also represented in the factor space of PRF scales. Figure 1 illustrates the extent of agreement obtained between the inferential and empirical PRF dimension. This implies that there is a veridicality to implicit personality theory in the sense that inferred relationships are mirrored in the real world. A particularly striking further application of the use of implicit personality theories is in a study by Stricker, Jacobs, and Kogan (1970). These authors discovered that female high school students, when asked to group personality items from the MMPI *Pd* scale on the basis of what they thought was being measured and then to name the trait or characteristics being measured, showed

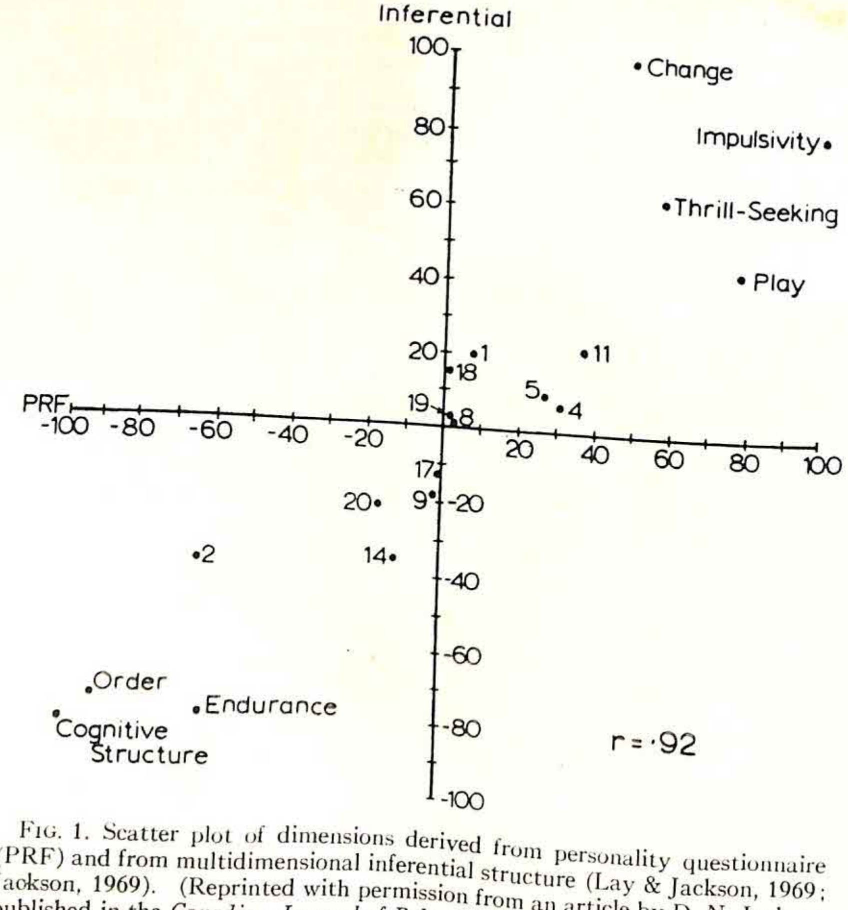


FIG. 1. Scatter plot of dimensions derived from personality questionnaire (PRF) and from multidimensional inferential structure (Lay & Jackson, 1969; Jackson, 1969). (Reprinted with permission from an article by D. N. Jackson published in the *Canadian Journal of Behavioural Science*, 1969, Vol. 1. Copyrighted by the Canadian Psychological Association, 1969.)

a remarkable congruence with factor-analytic investigations of the same content.

SAMPLING THE CONTENT DOMAIN AND THE PROBLEM OF SUBTLETY

What implications can be derived from results such as those described in the above section? It is now possible to state two important further principles which might be considered as corollaries of the first general principle regarding the theoretical foundations of personality assessment.

An item pool may be viewed as a sample of content; it should be evaluated in terms of how adequately and how representatively it reflects the content domain or universe of content implied by the definition. The question of sampling content hardly needs further elaboration except perhaps to suggest that judgment plays a very important role here. In fact it has no equal. Techniques which formalize human judgment, such as those employed in the multidimensional scaling studies described previously, provide a substantive basis for evaluating the possibility of discriminating between related constructs. If judges cannot discriminate between related constructs using such powerful techniques, it is quite unlikely that respondents can either. But at the initial stages of personality scale development, one is not likely to have items in sufficient quantity and quality with which to work. It is necessary, therefore, to define what one is seeking to measure and to develop an item pool.

Consider the dimension of risk taking. Here, one might wish to survey the variety of situations in which a person might undertake risks of various sorts. One would then establish a matrix of various situations involving risk, including those related to interpersonal relations, financial activities, health, bodily injury, and some of the assumed manifestations of risk taking in thrill seeking and need for variety. In addition, one would proceed to define the modes of behavior entailing risks of these various sorts and of typical risk-related activities, such as driving an automobile, working with machinery, or participating in sports. If it is assumed that risk taking is a more or less unitary dimension, one would expect manifestations of risk taking transcending each of these situations and showing a pattern of consistencies. Psychologists, prone to emphasize situational aspects of behavior, have pointed out that situations often determine the responses of groups of individuals. However, if a predominant response across all of the situations can be specified, and if these can be shown to be substantially associated with a single continuum, then one can speak of risk taking as a unitary dimension. The question of the evaluation of the viability of a trait or construct conception of personality characteristics thus remains empirical, employing the criteria of response homogeneity and stability, and heteromethod convergent and discriminant validity.

But, while recognizing the importance of these evaluative criteria, the prior question of the adequacy of item sampling remains. Short of actually sampling the behavior of very large numbers of individuals in their natural state in the manner of Barker's (1955) *One Boy's Day* magnified by a factor of 10,000, there is probably no way to judge the adequacy of an item pool and of a definition of a trait except by the use of rational processes, either of an individual psychologist or of a group of judges.

But if anyone wishes to dispute the utility of such rational processes in item selection, and to question whether such approaches involve ecological as well as construct validity, the author issues a challenge. For any trait for which substantive definition is possible, let the most elaborate empirical item-selection procedures using criterion groups be pitted against two hours of work by a couple of good item writers. In this case, it is suggested that the selection of item writers be substituted for the selection of items. If anyone wishes to accept the challenge, these two alternatives might well be appraised. The rules of the game would be that the empirical procedure would employ a heterogeneous item pool like that contained in the California Personality Inventory or the

MMPI, either published or unpublished, whereas the substantive approach would involve two item writers working for two hours each. It would be necessary to exclude item pools which already contain substantive item selection of traits similar to the one at issue. One might extend this challenge even further. It might even be possible to use unselected item writers. It might be interesting, for example, to have an introductory class of psychology students write one item each with regard to a defined dimension, with perhaps just a bit of screening for substantive cogency and clarity of style, and conduct the comparison on that basis. The comparison proposed would be, of course, that of the empirical validity against a criterion relevant to the construct in question. The author would fully expect under cross-validation that even an inexperienced item writer would be superior to empirical item selection with a typical heterogeneous item pool.

This, of course, is attacking the problem on the home ground of the empirical approach to item selection. This is the raison d'etre of this approach—all other considerations are secondary to empirical validity. The primary reason for believing in the superiority of substantive over empirical procedures is based on experience with the PRF and with a number of studies conducted with my associates at the University of Western Ontario, and by others (e.g., Goldberg & Slovick, 1967). In the case of the PRF, for example, none of the item-selection procedures had reference to the empirical validity of an item. However, when the correlation between each scale and a predicted peer behavior rating was appraised, the median validity proved to be .52 for 20 scales. A search of the literature has failed to uncover any instances of higher validities from any approaches focusing on empirical validity in item selection.

The reasons for the modest cross-validational records for empirical techniques are apparent. Rarely do item pools in use by those who seek to maximize empirical validity in item selection constitute an adequate sample of the behaviors relevant to a particular trait. Rarely are they sufficiently free from forms of response bias. Rarely are the criterion groups that well defined that they have not incorporated bias in their selection, and rarely do they constitute a comprehensive characterization of the trait. Rarely are scales so constituted mutually exclusive with discriminant properties. But if a rational item pool is developed carefully and well, these problems can be overcome; so too can the problem of item transparency in personality assessment, an observation which leads to a second corollary of the first principle.

Since good subtle items require a sensitive appreciation for the remote, discriminating manifestations of a trait, a necessary precondition for developing a pool of subtle items of adequate quality and quantity is substantive specification of item content. Of course, subtlety is a positive attribute in a personality questionnaire, particularly when one is measuring undesirable traits, such as in the previous example involving sadistic impulses. It might appear as a contradiction that both the direct representation of content as a sample of an appropriate universe and the positive values of subtlety are considered positive virtues. An example will help clarify this apparent paradox. Consider the item, "I think newborn babies look very much like little monkeys." A respondent might be hard pressed to tell what is being measured by such an item. But if he knew that this item was written to represent an attitudinal exemplar of the negative pole of the Nurturance dimension, and in fact did so very well, the interpretation of the item becomes clearer. Given a definition of a trait, it is possible to identify meaningful, valid, and relatively remote exemplars. They become remote by virtue of the fact that the individual may not know precisely what traits are being measured. But if the average judge knows about the nature of the trait being assessed, he can, as the social perception studies indicate, quite accurately infer the nature of the content, and we might be quite suspicious of it if he could not. This ability is not uniformly present in the same degree in all respondents (Stricker, 1969) but, like knowledge of connotative attributes of content, is differentially present (Jackson & Messick, 1969). Thus, there is a fundamental distinction between subtle items as described here and as identified in the tradition of empirical item selection. Whereas subtle items on the MMPI were identified post hoc solely on the basis of their sample estimate of a correlation with a criterion, we require, in addition to an empirical relationship, a substantive a priori link to an underlying construct.

Whereas Meehl argued for a projective facet to an item in which validity hinged on the *misinterpretation* of some aspect of the item, one might well argue for the opposite. Those items are the best which are clearest and which contain referents about which people agree. This is not to say that respondents have to see the connection between the underlying trait and the item. But in the example of the item about newborn babies looking like monkeys, it is reasonable to assume that individuals have had sufficient experience, either directly or indirectly, with both monkeys and babies to answer meaningfully. In the perhaps 15,000 items that the author has had occasion to review in the past decade, more than a few ambiguous items have slipped by his editorial screening. These almost invariably have proved to be items with mediocre statistical properties.

CHOICE OF A MEASUREMENT MODEL

The method of choice in personality scale construction is one which focuses on a linear relation between items and a single underlying latent continuum. Other approaches, embodying more complicated models such as dynamic and class models, have the status of curiosities. Much has been written in the last quarter century about the problem of measuring personality. Many have argued that personality is complex and that the linear approach of factor analysis and classical test theory is unequal to the task. Thus, there has emerged a variety of alternative models which, implicitly or explicitly, has guided test construction. Among these are the class model of psychiatric diagnosis implicit in the development of the MMPI, in which an individual is considered to be a member of one of several postulated classes, possibly genetically determined (cf. Meehl, 1959). Loevinger (1957) has called attention to dynamic models in which opposites appear similar, as in the manifestations of the psychoanalytic defense mechanism of reaction formation. There are the familiar arguments regarding configural scoring (Meehl, 1950) and of interpreting patterns of scores on a psychometric instrument.

Unfortunately, authors who argue for configural approaches rarely have taken the trouble (Meehl is an exception) to specify mathematically the model appropriate for their evaluation. Even more rarely has empirical evidence been adduced to support such conjecture. While the author's credentials, both in terms of a fairly extensive reading, for example, of psychoanalytic literature, and an early exposure to the lore surrounding the interpretation of the MMPI profiles, might encourage a sympathetic appraisal of such approaches, convincing shreds of empirical evidence for the preferability of a class, dynamic, or nonlinear model for combining items, or for any significant gain in applying configural scoring over the usual linear scoring to items, are encountered rarely, if at all.[5] This should remain an open question, and those willing to expend effort and energy in pursuing these areas deserve our warm, if skeptical, encouragement. But until really convincing evidence is forthcoming, the substantial body of knowledge in classical test theory should not be abandoned. Until that time, such models have the status of curiosities; their existence in isolated cases is not denied, but their low base rates do not encourage the expenditure of much time in their identification.

Jane Loevinger has taken an opposite point of view. She has been an incisive critic, not only of ad hoc methods of personality scale construction, but of classical test theory and factor analysis as well. One might hope, then, that her own investigations would provide more sensitive, subtle insights into personality than those provided by some earlier investigators. Unfortunately, the empirical work is at best discouraging. In one series of studies, Loevinger and her associate (Ernhart & Loevinger, 1969) used a type of homogeneity clustering to identify subsets of items empirically sharing overlapping variance. Five clusters of authoritarian family ideology were isolated, but on cross-replication all but one of these failed to reappear, and the one that did reemerge appeared to reflect a broad dimension highly correlated with intelligence, social class, and education, recalling earlier findings with the California F Scale. The F Scale has been shown to elicit a broad response-bias dimension, uncorrelated with attitudinal content, but related to style of wording (Clayton & Jackson, 1961), and substantially related to intellectual ability, educational level, and social class.

---

[5] A distinction should be made here between the configural scoring of *items*, to which the above comments apply, and the configural combination of scales. The latter approach may well prove to be a fruitful avenue to the investigation of psychopathological syndromes.

While Loevinger and her colleagues are to be commended for seeking cross-replication in a diversity of samples, the application of such a homogeneity keying procedure, without adequate concern for content sampling within scales or for response biases elicited by item style, involves something equivalent to the crudest application of a linear model. This results in total scores grossly oversimplifying complex combinations of variance from a number of sources. A procedure in which the dimensions to be measured are specified in terms of their putative manifestations, and in which the expected response biases are carefully controlled, would have been preferable. While the routine application of factor analysis may have certain problems, until its critics can devise more sensitive procedures, one cannot be impressed by the substitution of simplistic, atheoretical procedures for combining items.

PROBLEM OF RESPONSE STYLES

Now, if sufficient concern has been conveyed for measuring content, consideration may be directed to the fifth general principle regarding the dynamics of structured personality inventories. It concerns response styles, which may be considered complementary to the content component of response consistency. The principle is that *to construct psychological measures in disregard for sources of method variance is to court disaster.* It is to the credit of Meehl and Hathaway (1946) that they recognized that scales constructed on the basis of internal consistency analysis typically would emerge with broad dimensions of test-taking attitudes. Unfortunately, they did not carry the implications of this argument far enough. Had they done so, further attempts at scale construction might have proceeded quite differently and might have developed sets of scales far less saturated with broad general factors lacking in discriminant validity. One such issue is the definition of content and of style and the relative importance of each in particular assessment contexts. Jackson and Messick (1962) factor analyzed the MMPI on three separate samples, identifying two large dimensions, accounting for about 75% of the common variance. One factor was highly correlated with mean desirability scale values of scales, and the second separated true- and false-keyed subscales. Jackson and Messick interpreted these dimensions as stylistic (Jackson & Messick, 1958) in the sense that they were general over a wide range of content, were elicited by aspects of item form, that is, the desirability and direction of keying of items, and possibly reflected individual differences relevant to some criteria, such as the separation of normal from deviant populations. Block (1965) (cf. Block, 1967; Jackson, 1967a, 1967b) preferred to focus on a substantive interpretation of the two large MMPI factors, arguing that their criterion correlations implied interpretations in terms of "ego resilience," and "ego control." Rorer and Goldberg (1965) reported correlations between original and reversed MMPI scales which appeared to reflect response consistency to item content. However, Jackson and Messick (1965) identified the same two large dimensions in the

Rorer-Goldberg data—one interpreted as desirability and the second associated with the keying of scales. Most of the variance attributed to content in correlations between individual MMPI scales by Rorer and Goldberg proved to be a function of these two large factors, which Jackson and Messick interpreted as stylistic.

Obviously, the point of view reflected in the work of Jackson and Messick involves a different definition of content and of style than does that of proponents of the tradition of empirical item selection. The interpretation of the role of content and of style must inevitably reflect the definitions afforded each. One problem with the tradition of empirical keying with respect to external criteria is that its usual application affords no firm basis for imputing latent attributes to respondents as a function of their test scores, no firm basis for relating item content to scale content, no substantive criterion for the relevance of items to putative theoretical dimensions, no structural criterion for the relation of the pattern of item responses to the pattern of scale responses, and, finally, no necessary requirement of convergent and discriminant properties of scale scores (Campbell & Fiske, 1959) with respect either to the internal structure of the test or to external criteria. If one's definition explicitly incorporates such requirements (cf. Bentler, Jackson, & Messick, 1971; Jackson & Messick, 1965; Loevinger, 1957), then one would look with disfavor on imputing discriminant content characteristics to sets of highly redundant scales whose psychometric properties can be duplicated rather well with sets of items selected randomly with respect to specific content but systematically with respect to item form (Edwards, 1966; Jackson & Messick, 1962). On the other hand, if external validity is the sole criterion for including or excluding an item from a scale, the basis for consistent responses to items becomes secondary to its external validity. Thus, within this framework it really makes very little difference theoretically or practically whether one prefers the interpretation of extremely general "content" dimensions which fail to show substantial discriminant validity, or the interpretation of pervasive stylistic dimensions which show certain empirical correlates. If, in the tradition of radical empiricism, one eschews theoretical or substantive definition and prefers to rest one's case solely with external correlates, there is, indeed, no basis for distinguishing content from stylistic variance, and the issue becomes equivocal and a source of much needless contention. Carried to its extreme, this point of view places the investigator in the nihilistic position of denying the relevance of evidence relating to the response process, and hence being unable, apart from external validity, of specifying the nature of responses to content and to item form. While not denying the relevance of external validity, particularly convergent and discriminant validity, the author takes the position that a great deal is to be gained by learning something about why a person answers a particular item the way he does. To capture elusive content variance, it is essential to attend to all sources of variance,

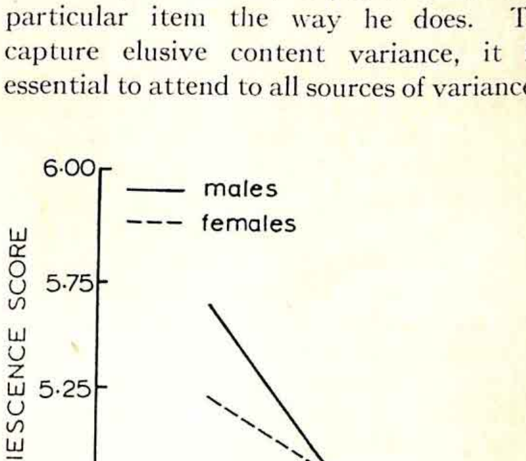


FIG. 2. Effect of speed of presentation upon acquiescence. (Reprinted with permission from an article by D. M. Trott and D. N. Jackson published in the *Journal of Experimental Research in Personality*, 1967, Vol. 2. Copyrighted by the Academic Press, 1967.)

including stylistic or method variance (Wiggins, 1968).

This viewpoint is reflected in a series of studies undertaken by the author with a number of collaborators at the University of Western Ontario. Trott and Jackson (1967) sought systematically to manipulate the parameters affecting acquiescence by varying item length, content dimensions, exposure time, content saturation, and subjects' verbal ability in two factorial designs each conducted separately for males and females. By selecting items so that content was balanced as to direction of keying, it was possible to obtain a score reflecting content, based on equal numbers of true- and false-keyed responses, and a second reflecting acquiescence, based on an all true key for each content scale. Of particular interest was the powerful effect of speed of presentation on acquiescence. Figure 2 shows that acquiescence decreased monotonically as a function of increasing presentation intervals. On a separate sample, when content scores were factor analyzed, the largest factor for the fast speed condition was one separating true- and false-keyed subscales for the same content dimension; for the longer exposure condition, a similar factor was only the fifth largest. Item length singly, and in interaction with speed of presentation, was significant in affecting acquiescence scores for males but not for college females (Figure 3). Content saturation, as reflected in separately determined biserial correlations with the total scale, interacted with both scale content and item length in affecting acquiescence. Trott and Jackson employed an information-processing rationale for these and similar findings. The critical role of content saturation was also highlighted in a pair of studies (Jackson & Lay, 1968; Morf & Jackson, in press). In the first study, separate factors were defined by marker variables for desirability, acquiescence, and adjective endorsement, but no substantial loadings were obtained on response-style factors for highly saturated PRF items, which defined only specific content dimensions. However, when Morf and Jackson selected PRF items which, although showing highest correlations with their own scale, did not correlate highly enough to be included in the PRF, every PRF subscale had a loading on at least one response style factor and on one content factor. In addition to stylistic factors marked by adjective checklist scales and PRF Desirability scales, Morf and Jackson identified two factors representing distinct acquiescence processes. The first, identified as *true responding*, represented, for example, double endorsement of both original and of reversed items from the same scale; the second, identified as *item endorsement*, involved the tendency to endorse items as self-descriptive regardless of direction of wording, as in responding true to a statement affirming the presence of a particular trait and false to a statement negating the presence of the trait. When these two factors were represented on orthogonal axes, 43 of the 44 scales fell in the predicted quadrant, a finding hardly to be expected by chance.[6] Indeed, the

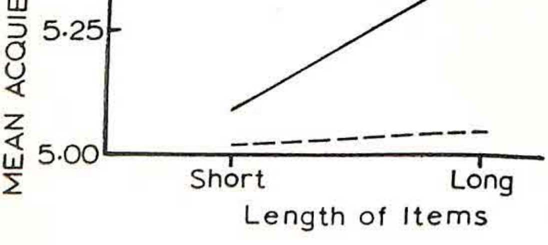


FIG. 3. Effect of item length upon acquiescence. (Reprinted with permission from an article by D. M. Trott and D. N. Jackson published in the *Journal of Experimental Research in Personality*, 1967, Vol. 2. Copyrighted by the Academic Press, 1967.)

---

[6] Block (1971) has recently conjectured that the results from the Morf and Jackson study are due to capitalization on chance in the rotation of axes. This was tested by Jackson and Morf by correlating component scores generated by two separate factor analyses of two distinct parallel sets of the variables based on the original data. The same was done for random data. Whereas for the real data component

regularity of these and similar findings may soon permit mathematical formulation of a theoretical model (cf. Jackson, 1968; Rogers, 1970). It is important to note, however, that four very clear content factors emerged in the Morf and Jackson analysis representing the four content scale types included in the analysis. Thus, there were eight identifiable factors, four having to do with content and four with response style. The point of this analysis was not to prove that either content or response style was predominant but rather to show that by a properly designed study, one can identify each.

There are, of course, other response styles in addition to the four identified by Morf and Jackson. There is, for example, the well-established defensiveness dimension identifiable in terms of MMPI Lie-scale-type items. Here, two item properties, desirability and frequency of endorsement (Jackson & Messick, 1969), operate in such a way that when there is a discrepancy between these two, it is inferred that the person is "lying." But evidence developed by Crowne and Marlow (1964), Kogan and Wallach (1964), and others would place a more sophisticated construct interpretation on this response style.

A number of years ago, Sechrest and Jackson (1963) spoke of various forms of deviation, and one form that struck them as potentially interesting for investigation was sheer randomness of responding. In constructing the PRF, a variety of items were assembled which, while not written to be bizarre, involved quite atypical behavior, such as "I learned to repair watches in Switzerland," or "I visited the Republic of Samoa during the past year." Subjecting these to item analyses, it was found that although their endorsement frequency was very low, they tended to show the convergent and discriminant properties expected of a content scale. Later, Bentler (1964) investigated this response style in predicting response stability when people were tested on different occasions, and he found that it accounted for approximately 50% of response variability for content scales. Bentler and Jackson (Jackson, 1970) also found that the test–retest reliability of a content scale proved to be an inverse function of the level of Infrequency scale score (Figure 4). This suggests that not only should one specify the reliability of a test, but also the reliability of samples of people, as relative reliability appears to be a stable property of people, as well as of scales (cf. Hendel & Weiss, 1970).

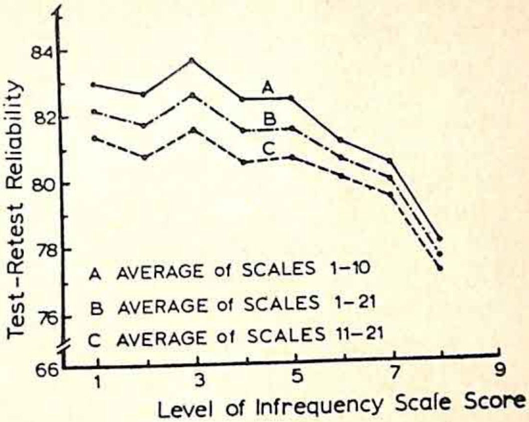


FIG. 4. Mean PRF scale retest stability as a function of Infrequency scale level.

Rather than provide a complete catalog of response styles, it is hoped that by describing a few, one should avoid the tempting alternative of proceeding by ignoring, ostrichlike, the possible impact of stylistic variance. Although the response-style literature has become rather technical in the psychometric sense, the personality research worker need not be discouraged. Fortunately, the evidence pointing to the strong effect of content saturation in reducing the impact of response style suggests that substantially defined scales which are purified by item-selection strategies seeking to maximize saturation can effectively suppress the role of response style. The omnipresence of evaluative bias virtually requires that one of several statistical procedures (Neill & Jackson, 1970) for identifying items implicated in this form of bias be incorporated into any serious attempt at scale construction. Item and

---

scores representing each of the factors assessed correlated at well beyond the .0001 level, none of the scores based on random data correlated significantly. Clearly, the null hypothesis may be rejected.

scale variance has finite limits; that accounted for by content saturation can powerfully reduce that traceable to stylistic factors. On the other hand, procedures such as empirical item selection, which are indifferent to content saturation, usually result in heterogeneous aggregates. When such heterogeneous items are summed, they yield scales reflecting largely what they have in common, namely, stylistic variance. Such scales fail to show really worthwhile levels of validity and are indistinguishable from many other similarly developed scales bearing quite different names and conceived for quite different purposes.

Multivariate Analysis of Structure

Perhaps it is appropriate after the discussion of response style to introduce a further principle. *A componential, multivariate approach to assessment is essential to the study of the structure of scales in order to develop the properties of convergent and discriminant validity.* One inevitably must consider the relationship between the manifest responses and the latent attribute that is being measured. In other words, if one wishes to infer something about an underlying trait, it is important to relate how a person responds to an item to the presence of the trait. As indicated above, there are a number of possible alternative models, including dynamic models and the classical univariate test theory approach. There are sophisticated approaches to homogeneity such as those outlined by Guttman (1950) and most recently by Bentler (1969). Univariate models generally will not stand by themselves in the analysis of the heterogeneous item pool. If one important aspect of test scores involves the multiplicity of response styles that have already been described, quite apart from content, a mindless search for consistency among sets of heterogeneous items will typically yield evidence of the tendency to respond in a particular stylistic way, contributing spurious forms of homogeneity and fulfilling the Meehl and Hathaway prophecy. This is not to say that homogeneity is not important, but it requires a more sophisticated and more delicate balance of content and the balancing of response styles in such a way that they do not importantly determine the structure of the total score. What is required is a careful specification of what is important to measure and what is not. If one is interested in measuring a single trait, both what the trait is and what it is not must be specified. A trait involving, for example, need for Social Recognition is different from one involving Exhibition in certain specific ways. If the conceptions underlying these definitions are accurate, these distinctions should be borne out in the structure of the item relationships and in the discriminant validity which will appear later. If the definitions of the traits are fuzzy or incorrect, these will also be reflected in item relationships. Similarly, if no clear distinction is made between a trait and a generalized tendency to respond undesirably, then it is likely that a number of such traits will show spuriously high intercorrelations and be inadequately differentiated. It is easy to make the error implicit in the multiple-correlation approach. Some investigators, for example, have argued that because acquiescence and attitude content are correlated with a particular criterion, they are, therefore, justifiably incorporated within the same scale. But if one is interested in measures which reflect constructs, the indiscriminate attempt to blend in unknown proportions of components of content and of style inevitably will lead to excessively highly correlated scales, replete with method variance and showing inadequate levels of discriminant validity (Jackson & Messick, 1962).

Much can be learned from an analysis of structure through the use of factor analysis. Factor analysis, however, may lead to an inappropriate focusing on characteristics either too broad or too narrow to be useful as explanatory constructs. Factor analysis at the item level often results in extremely specific, sometimes even trivial, exemplars of a trait. These trivial exemplars may or may not have general import in the description of per-

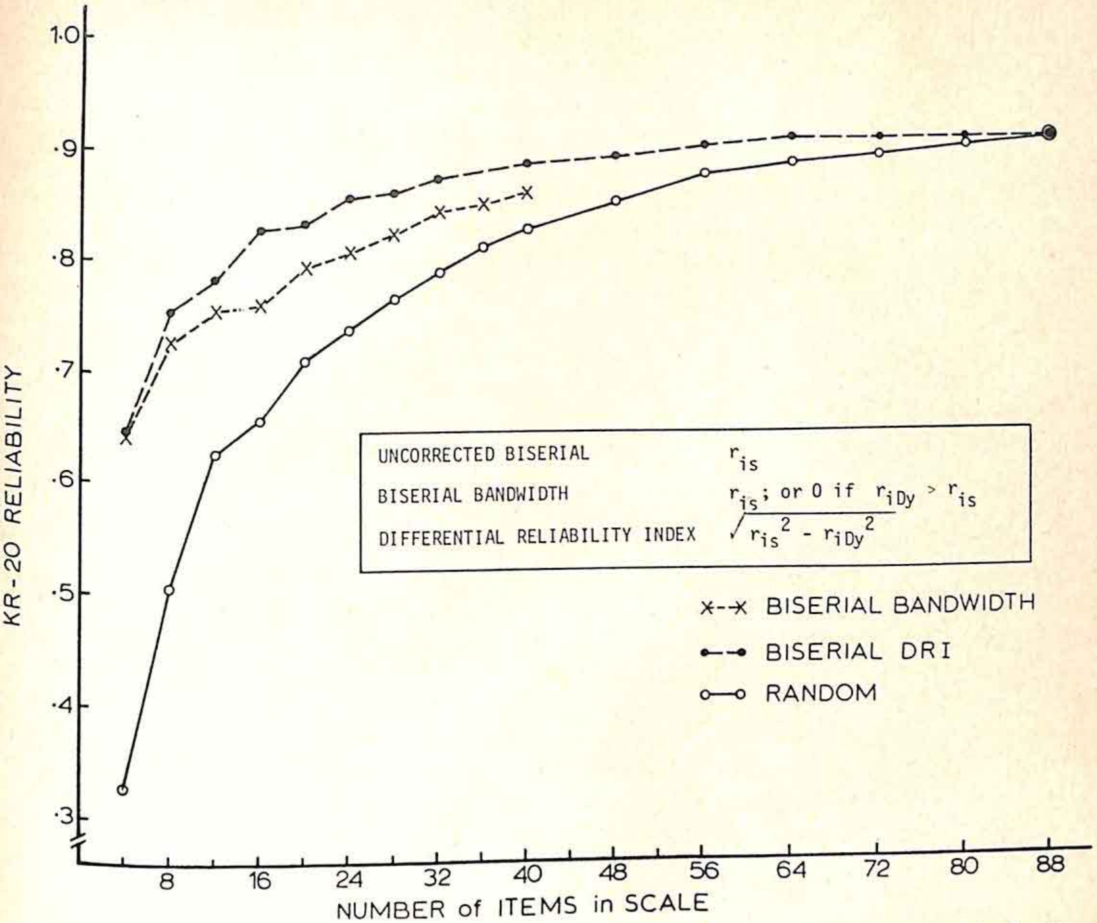


FIG. 5. Personality scale reliability as a function of number of items selected on the basis of three strategies of personality scale construction. (r_{is} = biserial correlation between item *i* and the total Evaluation Sensitivity scale; r_{iDy} = biserial correlation between Item *i* and the Desirability scale. (Reprinted with permission from an article by J. A. Neill and D. N. Jackson published in *Educational and Psychological Measurement*, 1970, Vol. 30, Copyrighted by Frederic Kuder, 1970.)

sonality, depending on whether or not important characteristics are represented in the item pool. At the item level, it has been shown psychometrically that there are purely mathematical relationships between the biserial and point-biserial correlation and the first centroid factor (Henrysson, 1962), leading to a preference for the computationally simpler approach of using classical item indexes of internal consistency. It is better, however, to incorporate a means for taking the massive desirability dimension into account. A number of alternative approaches are possible. Recently, Neill and Jackson (1970) completed a study in which more than a dozen indexes of item selection were contrasted, several of which used unique rationales for suppressing the desirability factor. Figure 5 presents a comparison of two of the indexes with a random strategy for selecting from a predefined substantive item pool in terms of reliability on a cross-replicated sample. The important conclusion was that, as with the PRF, it was possible very substantially to suppress the desirability dimension at the item level, and yet obtain high levels of content saturation with relatively short scales. Since that time, it has been possible to vastly elaborate item-selection procedures in a number of directions. Rather than use, for example, the total score as the criterion against which to select items, one may identify some component of the total score which is entirely independent and uncorrelated with unwanted sources of variance, such as the desirability dimension. It has also been possible to seek to maximize item variance and to seek to obtain maximum discrimination among related scales.

This latter point raises a particularly interesting question. Factor analysis, in the tradition of Thurstone and Cattell, has as one of its major alleged advantages the imposition of a more parsimonious interpretation among a set of statistical variables. In the case of personality variables, the hope is that one can reduce the rank of a correlation matrix in such a way that the entire matrix can be approximated by a matrix of smaller rank. Factor-analytic investigators have traditionally employed only a single method of measurement and have sought to identify the structure of personality among a set of personality scales. They might find, for example, that scales measuring Achievement and Endurance define one pole of a dimension and a scale reflecting Play defines the opposite pole. The personality theorist might object that these variables have a good deal of specificity which might be useful in measurement. But it is an element of strong faith on the part of the factor-analytic investigators that the truth is to be identified in what is common among variables, rather than what is unique. Inevitably, however, content and method variance within a single method of measurement usually have a similar structure (Campbell & O'Connell, 1967; Jackson, 1969b). This causes variables to correlate much more highly within a given method of measurement than they should, resulting typically in distorted factors often misrepresenting seriously the nature of the structure of psychological variables. This led to a need for developing procedures which would yield many, rather than few, factors and would focus on heteromethod validities. There are both sound psychological and psychometric reasons for preferring factor-analytic solutions which allow appropriate distinctions between method and trait factors and between theoretically distinct traits, even if this requires the sacrifice of maximum parsimony. For example, applying multimethod factor analysis (Jackson, 1969b) to PRF parallel forms, precisely 22 factors emerged, one for each trait, with only the appropriate pair of scales salient on each factor. In a similar manner, it was possible to show that 18 factors accounted for 20 traits in a space defined by PRF scales, peer behavior ratings, and self-ratings (Jackson & Guthrie, 1968).

It is possible to carry this line of reasoning considerably further and to suggest that factor-analytic investigators who have sought to identify reality within the structure of correlations between tests might better be advised to seek it in the reliable but specific variance within scales, keeping in mind the problems of method variance. Clearly, if one wishes to maximize the predictability of a battery, entirely uncorrelated tests would be appropriate (cf. Guttman, 1958). We urgently need methods of analysis which will aid us in approaching this ideal. This, as well as theoretical distinctiveness, will foster the kinds of convergent and discriminant properties of tests which Campbell and Fiske (1959) have considered so important. With these requirements is mind, it is clear that efforts should be directed not only at the evaluation of tests, but just as properly at a program of test construction focusing at every state on convergent and discriminant validity.

OVERVIEW

What can be said then about the picture that has been drawn of the dynamics of personality assessment with the benefit of a quarter of a century of further experience. In one sense, the picture is more complicated than the simplistic and perhaps mechanical model proposed in an earlier era. Personality items may be conceived of as eliciting a combination of a number of sources of variance including the following: content variance appropriate to the trait or disposition being measured, content variance associated with irrelevant dimensions, method variance, specific variance unique to the item, stylistic variance, and various interactions between stylistic and content variance in which, for example, content is highlighted by a particular format, or vice versa. Items can thus be viewed as differing in content saturation and in their stylistic-eliciting potential, while persons can be conceived of as differing in sensitivity to item content and to stylistic features, as well as showing varying positions on latent dimensions.

On the other hand, however, the picture may be considered simplified. One need not abandon as unscientific his unique capacity to judge and weigh content. One need not go out and seek a new properly sampled criterion group or groups every time there is a very particular criterion to predict, nor does one need to change personality scales, like style in clothing, every year or two, as might be the case with empirically derived scales. While in one sense a response to a personality item bears the unique stamp of the person, it does not do so with equal precision for all of his traits. The Pandora's Box of blind empiricism in which a single item is keyed on a number of scales merely because it has correlated with different criteria can be rejected as both inefficient and troublesome. While responses to personality items are multifaceted, as are the items themselves, there is no reason to abandon completely conceptions of trait and test theory as they have evolved since the beginning of scientific psychology.

There is, indeed, reason to be satisfied with the enormous progress that has been made in the past quarter of a century in our understanding of the dynamics of personality inventories, progress which in no small part is due to the efforts of individuals like Paul Meehl, who proceeded optimistically forward when many confidently advised that it could not be done.

REFERENCES

BARKER, R. G., & WRIGHT, H. F. *One boy's day: A speciman record of behavior*. New York: Harper & Row, 1955.

BENTLER, P. M. Response variability: Fact or artifact? Unpublished doctoral dissertation, Stanford University, 1964.

BENTLER, P. M. An implicit metric for ordinal scales: Implications for assessment of cognitive growth. Paper presented at the Invitational Conference on Ordinal Scales of Development, Monterey, California, February 1969.

BENTLER, P. M., JACKSON, D. N., & MESSICK, S. Identification of content and style: A multidimensional interpretation. *Psychological Bulletin*, 1971, **76**, in press.

BLOCK, J. *The challenge of response sets.* New York: Appleton-Century-Crofts, 1965.

BLOCK, J. Remarks on Jackson's "review" of Block's *Challenge of response sets. Educational and Psychological Measurement*, 1967, **27**, 499–502.

BLOCK, J. On further conjectures regarding acquiescence. *Psychological Bulletin*, 1971, **76**, in press.

CAMPBELL, D. P. The development of the SVIB: 1927–1967. In P. McReynolds (Ed.), *Advances in psychological assessment.* Vol. 1. Palo Alto, Calif: Science & Behavior Books, 1968.

CAMPBELL, D. P. *Manual for the Strong Vocational Interest Blank.* Stanford, Calif.: Stanford University Press, 1969.

CAMPBELL, D. P., BORGEN, F. H., EASTES, S. H., JOHANSSON, C. G., & PETERSON, R. A. A set of basic interest scales for the Strong Vocational Interest Blank for men. *Journal of Applied Psychology*, 1968, **52** (6, Pt. 2).

CAMPBELL, D. T., & FISKE, D. W. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 1959, **56**, 81–105.

CAMPBELL, D. T., & O'CONNELL, E. J. Methods factors in multitrait-multimethod matrices: Multiplicative rather than additive? *Multivariate Behavioral Research*, 1967, **2**, 409–426.

CLAYTON, M. B., & JACKSON, D. N. Equivalence range, acquiescence, and overgeneralization. *Educational and Psychological Measurement*, 1961, **21**, 371–382.

CRONBACH, L. J. Response set and test validity. *Educational and Psychological Measurement*, 1946, **6**, 475–494.

CRONBACH, L. J. Further evidence on response sets and test design. *Educational and Psychological Measurement*, 1950, **10**, 3–31.

CRONBACH, L. J., & MEEHL, P. E. Construct validity in psychological tests. *Psychological Bulletin*, 1955, **52**, 281–302.

CROWNE, D. P., & MARLOWE, D. *The approval motive: Studies in evaluative dependence.* New York: Wiley, 1964.

EDWARDS, A. L. A comparison of 57 MMPI scales and 57 experimental scales matched with the MMPI scales in terms of item social desirability scales values and probabilities of endorsement. *Educational and Psychological Measurement*, 1966, **26**, 15–27.

ERNHART, C. B., & LOEVINGER, J. Authoritarian family ideology: A measure, its correlates, and its robustness. *Multivariate Behavioral Research Monographs*, 1969, No. 69–1, 3–62.

GOLDBERG, L. R., & SLOVIC, P. The importance of test item content: An analysis of a corollary of the deviation hypothesis. *Journal of Counseling Psychology*, 1967, **14**, 462–472.

Guttman, L. The basis for scalogram analysis. In S. Stouffer et al. (Eds.), *Measurement and prediction.* Princeton: Princeton University Press, 1950.

Guttman, L. To what extent can communalities reduce rank? *Psychometrika*, 1958, **23**, 297–308.

Hendel, D. D., & Weiss, D. J. Individual inconsistency and reliability of measurement. *Educational and Psychological Measurement*, 1970, **30**, 579–594.

Henrysson, S. The relation between factor loadings and biserial correlations in item analysis. *Psychometrika*, 1962, **27**, 419–424.

Jackson, D. N. Review of J. Block, *The challenge of response sets. Educational and Psychological Measurement*, 1967, **27**, 207–219. (a)

Jackson, D. N. Balanced scales, item overlap and the stables of Augeas. *Educational and Psychological Measurement*, 1967, **27**, 502–507. (b)

Jackson, D. N. *Personality Research Form manual.* Goshen, New York: Research Psychologists Press, 1967. (c)

Jackson, D. N. A threshold model for stylistic responding. In C. Hanley (Chm.), New interpretations of response style and content in personality assessment. Symposum presented at the meeting of the American Psychological Association, San Francisco, September 1968.

Jackson, D. N. The multidimensional nature of human social perception. *Canadian Journal of Behavioural Science*, 1969, **1**, 227–262. (a)

Jackson, D. N. Multimethod factor analysis in the evaluation of convergent and discriminant validity. *Psychological Bulletin*, 1969, **72**, 30–49. (b)

Jackson, D. N. A sequential system for personality scale development. In C. D. Spielberger (Ed.), *Current topics in clinical and community psychology.* Vol. 2. New York: Academic Press, 1970.

Jackson, D. N., & Guthrie, G. M. Multitrait-multimethod evaluation of the Personality Research Form. *Proceedings of the 76th Annual Convention of the American Psychological Association*, 1968, **3**, 177–178. (Summary)

Jackson, D. N., & Lay, C. H. Homogeneous dimensions of personality scale content. *Multivariate Behavioral Research*, 1968, **3**, 321–338.

Jackson, D. N., & Messick, S. Content and style in personality assessment. *Psychological Bulletin*, 1958, **55**, 243–252.

Jackson, D. N., & Messick, S. Response styles on the MMPI: Comparison of clinical and normal samples. *Journal of Abnormal and Social Psychology*, 1962, **65**, 285–299.

Jackson, D. N., & Messick, S. Acquiescence: The nonvanishing variance component. *American Psychologist*, 1965, **20**, 498. (Abstract)

Jackson, D. N., & Messick, S. A distinction between judgments of frequency and of desirability as determinants of response. *Educational and Psychological Measurement*, 1969, **29**, 273–293.

Kogan, N., & Wallach, M. A. *Risk taking.* New York: Holt, Rinehart & Winston, 1964.

Lay, C. H., & Jackson, D. N. Analysis of the generality of trait-inferential relationships. *Journal of Personality and Social Psychology*, 1969, **12**, 12–21.

Loevinger, J. Objective tests as instruments of psychological theory. *Psychological Reports*, 1957, **3**, 635–694.

Meehl, P. E. The dynamics of "structured" personality tests. *Journal of Clinical Psychology*, 1945, **1**, 296–303.

Meehl, P. E. Configural scoring. *Journal of Consulting Psychology*, 1950, **14**, 165–171.

Meehl, P. E. Some ruminations on the validation of clinical procedures. *Canadian Journal of Psychology*, 1959, **13**, 102–128.

Meehl, P. E., & Hathaway, S. R. The *K* factor as a suppressor variable in the MMPI. *Journal of Applied Psychology*, 1946, **30**, 525–564.

Morf, M. E., & Jackson, D. N. An analysis of two response styles: True responding and item endorsement. *Educational and Psychological Measurement*, in press.

Murray, H. A. *Explorations in personality.* Cambridge, Mass.: Oxford University Press, 1938.

Neill, J. A., & Jackson, D. N. An evaluation of item selection strategies in personality scale construction. *Educational and Psychological Measurement*, 1970, **30**, 647–661.

Rogers, T. B. Evaluation of a threshold theory for personality assessment. Unpublished doctoral dissertation, University of Western Ontario, 1970.

Rorer, L. G., & Goldberg, L. R. Acquiescence in the MMPI? *Educational and Psychological Measurement*, 1965, **25**, 801–817.

Sechrest, L., & Jackson, D. N. Deviant response tendencies: Their measurement and interpretation. *Educational and Psychological Measurement*, 1963, **23**, 33–53.

Stricker, L. J. "Test-wiseness" on personality scales. *Journal of Applied Psychology*, 1969, **53**, (3, Pt. 2).

Stricker, L. J., Jacobs, P., & Kogan, N. Veridicality of implicit personality theories. *Proceedings of the 78th Annual Convention of the American Psychological Association*, 1970, **5**, 157–158. (Summary)

Trott, D. M., & Jackson, D. N. An experimental analysis of acquiescence. *Journal of Experimental Research in Personality*, 1967, **2**, 278–288.

Wiener, D. N. Personality characteristics of selected disability groups. *Journal of Clinical Psychology*, 1948, **4**, 285–290.

Wiggins, J. S. Personality structure. In P. R. Farnsworth (Ed.), *Annual Review of Psychology*, 1968, **19**, 293–350.

(Received June 11, 1970; revision received January 6, 1971)

Psychological Review
1971, Vol. 78, No. 3, 249–259

DISCREPANCY HYPOTHESES:

METHODOLOGICAL AND THEORETICAL CONSIDERATIONS[1]

HOBEN THOMAS[2]

Pennsylvania State University

The conceptual framework of Coomb's data theory is used to highlight similarities and differences between two representative but different discrepancy hypotheses: McClelland's hypothesis and Dember and Earl's complexity discrepancy theory. It is concluded that mapping of data into the models cannot be done unambiguously unless assumptions are made concerning as yet unspecified parameter values and properties of data. Two different research strategies are discussed. One two-stage strategy with simplifying assumptions is suggested as a possible approach toward quantitative specification of model parameter values.

Considerable interest has recently been shown in discrepancy hypotheses as theoretical explanations for motivational processes in diverse content areas. These areas include intellectual development (e.g., Hunt, 1965), verbal processes (Kammann, 1966), exploratory behavior (e.g., Walker, 1964), and environmental psychology (Wohlwill, 1966). Discrepancy hypotheses share conceptual similarities with various inverted-U hypothesis formulations (e.g., Schroder, Driver, & Streufert, 1967; Vitz, 1966) which state that behavioral measures, often affective in nature, are nonmonotonic inverted-U functions of stimulation. Conceptually related to inverted-U hypothesis formulations is the notion that there is an optimal level of stimulation which is most pleasant or most effective in mediating behavior (e.g., Fiske & Maddi, 1961). Both of these general notions are very old and go back at least to Wundt in 1874 (Berlyne, 1960, p. 201). Discrepancy hypotheses are consistent with various inverted-U formulations, but they go further and tend to emphasize the behavioral effects of an optimal discrepancy or difference between some residual (often times momentary) organismic state and some source of stimulation. Early statements of discrepancy hypotheses which have been particularly influential include Hebb (1949), Dember and Earl (1957), and McClelland and Clark (1953). These statements remain as perhaps the most popular in the literature.

This paper attempts to clarify distinctions between discrepancy hypotheses and to specify clearly underlying assumptions together with parameters which ultimately must be specified if the models are to be developed further. To achieve these goals, two models have been selected for primary consideration, McClelland's hypothesis (McClelland & Clark, 1953) and Dember and Earl's (1957) complexity-discrepancy model. Emphasis is placed on these models primarily because they appear to be the most well-stated positions. However, it is argued that the models still possess too many unspecified free parameters for unambiguous mapping of data into the models, or for unambiguous predictions. Thus, whether data are said to support (e.g., Kammann, 1966) or fail to support (e.g., Verinis, Brandsma, & Cofer, 1968) a discrepancy hypothesis often boils down to differences in assumptions concerning the nature of data and selection of model parameter values. It is argued that there are equally plausible and perhaps more parsimonious alternative interpretations of such data.

---

[1] Some of the research reported was supported by Grant MH-04726 from the National Institute of Mental Health, United States Public Health Service, and the Central Fund for Research, Pennsylvania State University. The author is indebted to C. N. Cofer, C. H. Coombs, R. S. Lockhart, F. Wicker, and E. Willemsen for valuable comments on earlier drafts of this paper.

[2] Requests for reprints should be sent to Hoben Thomas, Department of Psychology, Pennsylvania State University, 417 Psychology Building, University Park, Pennsylvania 16802.

The terminology of data theory (Coombs, 1964) has been selected as a common conceptual framework within which to view similarities and differences between models. This paper begins with a consideration of the models as viewed within this framework after which difficulties of making unambiguous predictions from the models are considered. Most research relating to discrepancy hypotheses and other inverted-U hypothesis formulations is characterized as qualitative, because research strategies have been primarily concerned with finding some nonmonotonic inverted-U function (e.g., Haber, 1958; McCall & Kagan, 1967; Vitz, 1966). Problems of this approach and limitations of data interpretation are considered. Since qualitative procedures cannot be used to specify discrepancy model parameters in the absence of stronger, probably unrealistic assumptions, a two-stage procedure for quantitatively estimating parameters is described.

DISCREPANCY HYPOTHESES AS QIa DATA MODELS

As viewed here, the conceptual notion of discrepancy implies an operationally definable metric or distance. Specifically a discrepancy hypothesis is defined as some measurable real distance between an organismic state and some hypothetical optimally situated stimulus most effective in mediating behavior. Stimuli at other than an optimal distance from the organism have different behavioral effects. Thus, discrepancy hypotheses should be viewed as another class of models which attempt to represent behavior within a spatial framework. At present, however, theoretical formulations have been implicitly restricted to consideration of behaviors which may be represented in a single dimension. Discrepancies may be represented by distances along a line.

A common conceptual (and spatial) framework for considering the hypotheses is Coombs' (1964) theory. All forms of the discrepancy hypotheses appear to correspond with what Coombs calls QIa models for preference data. In QIa, subjects (*S*s) and stimuli are mapped into, and may be represented as, points within the same psychological space called a *joint space.* The basic notion is an *S* will order his preference for stimuli as some function of the *S*'s distances from the stimuli in this space.

Figure 1 illustrates three possible preference functions for the unidimensional case. Figures 1A and 1B represent the two different discrepancy hypotheses and illustrate that discrepancy hypotheses vary in stating how these preferences will be ordered as a function of *S*–stimulus interpoint distances. Figure 1C is considered later. In each figure, the abscissa is a *J* or *joint* scale of *S*s and stimuli. A, B, and C represent three hypothetical stimulus points with ordinates located at unequal distances along the *J* scale. Three stimuli are sufficient to illustrate the problems to be discussed. Following Coombs, the location of individuals (or groups of *S*s) will be called I or ideal locations. These locations may be thought of as adaptation levels. While stimulus points and I locations were selected for illustrative purposes, they are probably typical of experimental situations. In Figure 1, the heights of the stimulus ordinates represent preference. Preference is interpreted in a specific way. It represents a response by an *S* to stimuli where the stimuli are evaluated by being referenced to *S*'s own internal standard. A preference of steak over chicken or Camels over Winstons would be examples. Examples of preference measures used in studies to evaluate discrepancy hypotheses include, as measures of high preference, "looked at longest," "played with the most," "liked," while low preference would be, respectively, "looked at least," "avoided," "disliked" (e.g., Earl, 1957; Haber, 1958; McCall & Kagan, 1967). Preference represents the dependent variable which typically will vary from one content area to another. However, not all measures *called* preference are QIa data. For example, preferences for lots of money or high grades are probably evaluated with respect to an absolute external standard and are not preference data as viewed within data theory.

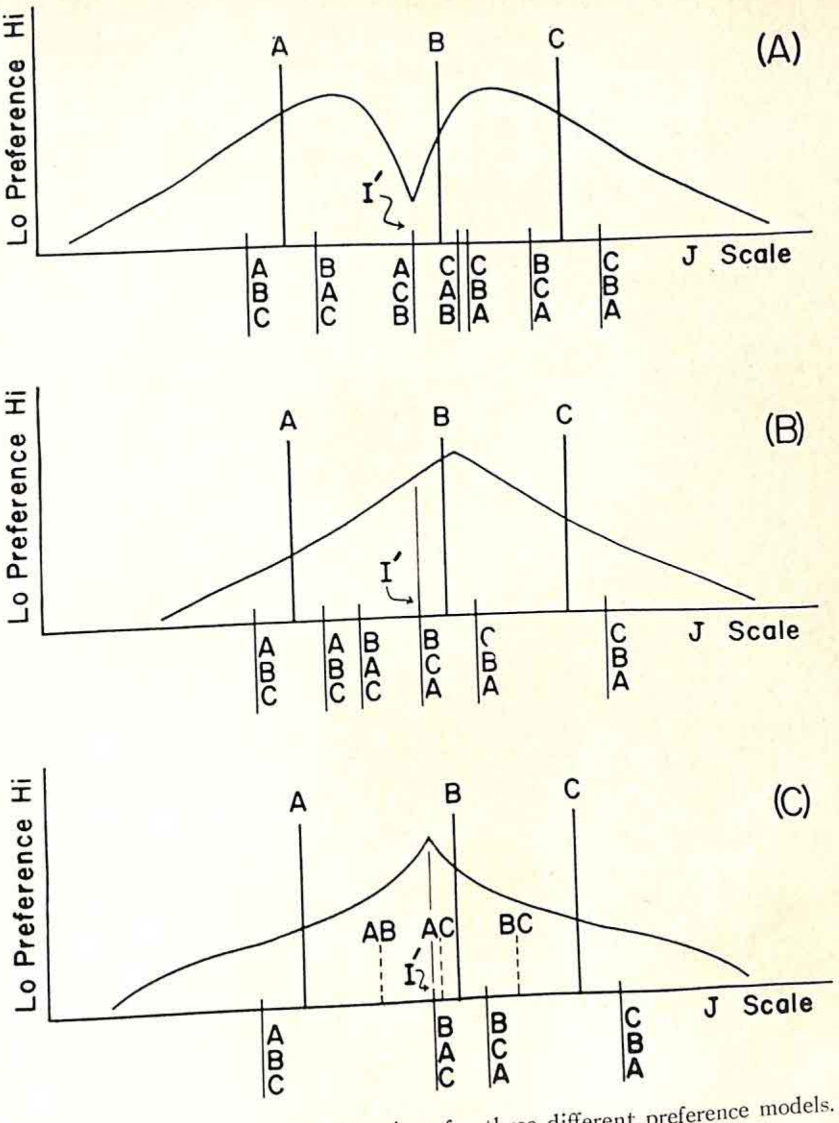


Fig. 1. *J*-scale preference functions for three different preference models. (A = McClelland's discrepancy hypothesis; B = Dember and Earl's complexity-discrepancy theory; C = Coombs' preference function.)

Figure 1A represents McClelland and Clark's curve relating preference for stimuli. The curve shown in Figure 1A is arbitrary but satisfies their conditions. It is symmetrical, nonmonotonic, an inverted double U curve (McClelland & Clark, 1953, p. 43). Relative preference is specified by the height of the ordinates. For *S*s located at I′, the preference order relation is ACB with A preferred most, and B least. This preference order relation is indicated directly below I′. While preserving the shape of the curve, it may be moved to the right or left such that the V or "trough" coincides with any one of an infinite number of possible *J*-scale locations. Stimulus points along the *J* scale remain fixed. From among many, six other I's have been selected and are indicated. These I's exhaust the different possible preference order relations, excluding ties. Thus, shifting the curve slightly to the right would produce preference order relation CAB, slightly further to the right CBA, etc. Slight changes in curve location do not necessarily produce different preference order relations because preference order relations are associated with discrete regions along the *J* scale.

Figure 1B represents Dember and Earl's (1957) preference function. Their formulation is of central importance to theories developed by Sackett (1965) and Walker (1964) and appears very similar to the theory suggested by Fowler (Fowler, 1967, p. 219). For Dember and Earl, the *J* scale is defined as a complexity scale. They predict that highest preference is for a slightly more complex "pacer" stimulus, with remaining stimuli being preferred in terms of their similarity or distance from the pacer. This implies a symmetrical curve peaking at the optimum hypothetical pacer with I some unknown distance below this peak. As illustrated in Figure 1B, the curve is nonmonotonic inverted U-shaped (or V-shaped) for stimuli more complex than individuals at I′ and below I′ or to the left, a continuously falling-off curve. Figure 1B is interpreted in the same way as Figure 1A. For I′ the preference order relation is BCA. Displacing the curve to the right or left but preserving the same peak-to-I′ distance produces different preference order relations as indicated. Figure 1A and 1B may be directly compared because five of the I locations are identical.

DISCREPANCY MODELS AND THEIR AMBIGUOUS DATA EXPECTATIONS

Before considering their predictive difficulties, it needs to be stressed that discrepancy models as currently developed assume that the responses to stimuli satisfy the criterion of a unidimensional ordered scale. The assumption is crucial. If it is not satisfied, comparisons among stimuli are difficult to make, since it could be maintained that *S*s were responding to the stimuli as if they were represented on different dimensions. In fact, it is argued below that this is a reasonable interpretation for many peculiarities in data. Thus, unidimensionality is a necessary prerequisite for interpreting data within discrepancy models. Yet, as far as is known, the assumption has not been treated as a testable hypothesis to be answered in data in any published study designed to evaluate any discrepancy hypothesis. The problem is particularly troublesome given the trend toward selection of heterogeneous stimuli such as paintings of modern art (Wohlwill, 1968) and poetry (Kammann, 1966). That responses to such multiattribute stimuli satisfy unidimensionality is an assumption which appears difficult to defend on rational grounds.

That which follows requires that the assumption of unidimensionality hold.

Regardless of methodology, that is, the dependent variable measured, or the experimental unit of interest (individuals or groups), or whether stimuli are presented singly or in sets, to compare data with a discrepancy model, the data must ultimately be reduced to preference order relations. For the present, it makes no difference whether the McClelland hypothesis is viewed with an eye toward predicting preference order relations, or whether the data are to be mapped into the model in a post hoc manner. The outcome is the same.

Figure 1A illustrates that under conditions of the McClelland hypothesis, if the I locations of the *S*s were unknown, or were expected to be different for different *S*s, any preference order relation is possible. Thus, all responses would fit the model unless there is some a priori way of knowing which preference order relations were more probable. Furthermore, *S*s could have different I locations and produce the same preference order relations. Thus CBA could be obtained from *S*s in two different noncontiguous regions of the *J* scale. Suppose however, as in some experimental procedure, the *S*s' I location could be specified. Let that location be I′. Then preference order relation ACB would be expected. But note this preference order relation is not invariant for this particular I location. If changes were made in the shape of the curve, or there were changes in the interpoint stimulus distances from those which were arbitrarily specified in Figure 1A (both changes can easily be visualized), then again it is not clear which preference order relations to expect. Not always, of course, would all possible preference order relations fit the model when many stimuli were used. The point to be emphasized, however, is that knowing which preference order relations to expect cannot be determined in the absence of other stimulus distance relations and model parameter values which have never been specified.

Consider next the Dember and Earl hypothesis, illustrated in Figure 1B. When I locations cannot be specified, the preference order relations compatible with the model are restricted, in the case of three stimuli, to four of the six possible preference order relations. Both CAB and ACB are incompatible with this model. This fact is considered below when the Coombs model is discussed. Here it may be noted that data expectations for single-peaked curves are easier to predict than those from double-peaked curves.

For preferences to the right of I′ in Figure 1A and 1B, no conceptual distinction can be made between the two models, and they share common problems. Comparing data expectations for the stimuli to the right of the I′ locations in Figure 1A and 1B illustrates difficulties when shapes are changed. In Figure 1A, at I′, C is preferred over B, but the reverse is true in Figure 1B. Furthermore, if the curve shape is preserved in Figure 1B but the I′-to-curve-peak (pacer) distance is changed, the preference order relation changes. This can be visualized by shifting the curve peak toward stimulus C, but keeping I′ fixed, until for individuals at I′, preference order relation CBA would be expected.

In summary, it can be stated that unambiguous mapping of data into discrepancy hypotheses or achieving unambiguous predictions from discrepancy hypotheses requires specification of the following parameters: (a) curve shape, (b) I location (or locations), and (c) distances between stimuli and between stimuli and I locations. Note that these parameters must be specified jointly. In the absence of parameter specification, almost any data may be mapped into or found at variance with the models by making different assumptions about parameter values. In addition to the above considerations, the unidimensionality assumption must hold.

QUALITATIVE TESTS OF DISCREPANCY HYPOTHESES

Lacking estimates of parameter values concerning distances along the *J* scale and the shape of the preference curve, studies testing discrepancy hypotheses have been restricted to predictions regarded here as qualitative, because the predictions concern only the general form of the preference function and not its specific shape. Actually, however, such predictions are not of central importance to *discrepancy* hypotheses.

For example in the case of McClelland's hypothesis, if an I can be specified it would be expected that preference above or below I should reveal progressive increases in preference until a maximum is reached after which preference should continuously fall off for stimuli located beyond this maximum. Such a procedure requires that I locations imbedded within a stimulus series be specified to within an order-preserving transformation, and that the stimuli to one or both sides of I be assumed to span a sufficiently broad range of the *J* scale such that if a "peak" preference exists, it will be revealed in the preference order relations. It would have to be assumed that the responses to the stimuli satisfied undimensionality. Since metrics along the *J* scale and the curve shape parameters are unknown, no specific preference order relation concerning which stimulus should be maximally preferred can reasonably be predicted, although such predictions have been made (e.g., McCall & Kagan, 1967). Hypotheses regarding distances from I to maximally motivating optimal discrepancies, hypotheses which are truly central to the models, would require that distance between stimulus points and between stimulus points and I locations be known.

Given the prerequisite conditions, the experimental procedure involves searching for some kth stimulus to one side of I for which preference is maximum. Preferences for stimuli on both sides of k must fall off strictly monotonically (toward I for stimuli to one side of k; away from I for stimuli on the other side of k). This condition defines a single peaked preference function. Let N_a equal the number of stimuli above I, and include the stimulus coincident with I if any. In the Appendix it is shown that the probability of finding a single peaked preference function is

$$p(A) = \frac{2^{(N_a-1)} - 2}{N_a!} \qquad [1]$$

The numerator is the number of such single peaked preference functions. The denominator is the number of possible preference order relations. The equation is useful when there are no ties in preference and for $N_a \geqslant 3$. It assumes error-free data, that is, no chance fluctuations in the function relating preference with stimuli. With moderate numbers of stimuli, the probability of finding a single peaked preference function given that the preference order relations were randomly generated is remote. For example, if $N_a = 6$, $p(A) < .05$. Equation 1 would hold for stimuli below I as well as for stimuli above I for McClelland's hypothesis. Let N_b equal the number of stimuli below I and include the stimulus coincident with I, if any. Assuming preferences above and below I are independent, if $p(A) \cdot p(B) < \alpha$, the McClelland hypothesis would have qualitative support. With N_a and $N_b \geqslant 3$, and the total number of stimuli > 7, $\alpha < .05$.

For Dember and Earl, Equation 1 would hold for stimuli above I, but there is only one preference order relation which falls off strictly monotonically below I; its chance probability is $1/N_b!$

Consider now studies which have made qualitative tests of discrepancy hypotheses. Haber (1958) adapted the hands of his *S*s to water temperatures, after which they made preference judgments for water at temperatures below and above their adaptation levels (i.e., I locations). Usually $N_a = N_b = 5$, and the form of the preference curves obtained for each *S* were in perfect harmony with the McClelland hypothesis. Haber's success can probably be traced to superior method considerations. His stimuli were psychologically simple; to control for judgment error he replicated both within and across *S*s, and Haber carefully assessed his *S*'s adaptation level before obtaining judgments.

Verinis et al. (1968) report four studies intended to provide data relative to McClelland's hypothesis. They conclude that each study fails to support it. The studies illustrate a number of problems regarding valid tests of discrepancy hypotheses in general, and interpretation of negative data in particular. For example, in their first experiment, college students read brief paragraphs concerning blind dates, college grades, Christmas job earnings, etc. These paragraphs were designed to establish an expectancy set (I). Afterward, *S*s were provided with information at different degrees of discrepancy from this set. The *S*s then evaluated on rating scales pleasures and motivation associated with the discrepant information. In general, monotone functions were found. The greater the positive discrepancy, the greater the pleasure. The greater the negative discrepancy, the greater the displeasure.

Given the nature of the stimuli, this finding would be expected. As viewed here, their stimuli called for stimulus comparisons (i.e., Coombs QIII data), not QI or preference data. It is hardly surprising that college students were unhappy about the prospects of making any lower quiz grade, or were most happy making a lot more, rather than just a little bit more, money on their Christmas jobs. Such judgments of money, grades, gas mileage, etc., are not preference data as viewed within data theory.

This difficulty aside, are the data incompatible with the McClelland hypothesis as claimed? Actually, by so interpreting the data, they are probably making unwarranted assumptions about unknown parameters. As stated, they found a monotone function falling off below I. Consider this finding in relation to Figure 1A. Assume that their *S*s were located at I′. Suppose that the stimuli Verinis et al. selected happened to have been located below I′ and to the left of the peak of maximum preference. Under these conditions, preference would be expected to be a decreasing monotone. By similar reasoning it can be seen that preferences might be an increasing monotone if the stimuli selected happened to be, for example, below I′ but to the right of the peak of maximum preference.

Thus, failing to find a single peaked preference function cannot be taken as evidence against inverted-U (or V) hypotheses unless strong, probably unrealistic assumptions are made concerning properties of data and model parameters. Studies which report negative data have made such assumptions (Block, 1964; McCall & Kagan, 1967; Verinis et al., 1968).

Except for Haber's (1958) study, data reported to be in correspondence with discrepancy hypotheses or other inverted-U hypotheses have typically shown more than one and often several peaks in the preference function (i.e., multiple peaks in the curve to one side of the individual's I location). Thus, the functions are often complex with many fluctuations (e.g., Conners, 1964; Kammann, 1966; Verinis et al., 1968). Two possible interpretations can be given for these curve fluctuations besides, of course, rejection of the discrepancy hypothesis under test. The interpretation which is implicitly accepted is that the multiple inflections in the curve are chance deviations from the "true" function. This interpretation is difficult to evaluate since there is no known error distribution—a troublesome problem, but one hardly unique to this situation. A second interpretation is that the responses were multidimensionally determined.

There are two reasons why this second interpretation is just as reasonable. First, the stimuli used in many studies are very complex multiattribute stimuli. As an illustrative example, consider Kammann's (1966) test of the Dember and Earl model. Among the findings were preference functions which were bimodal, a finding clearly incompatible with the model under test. But Kammann's stimuli were poems, which were probably responded to in no simple unidimensional manner. Other studies with similar difficulties include Conners (1964), McCall and Kagan (1967), Walker and Walker (1964), and Wohlwill (1968). A second reason for this interpretation is that Tversky and Russo (1969) have shown that when *S*s are *told* the stimulus dimension along which to respond, their judgments are still influenced by other stimulus properties. Thus, unless *S*s are constrained in their preferences, or data are tested to evaluate the degree to which a unidimensional criterion is satisfied, the interpretation of results when complex preference functions are obtained is unclear.

SUGGESTED PROCEDURE FOR QUANTITATIVELY ESTIMATING DISCREPANCY HYPOTHESIS PARAMETERS

An important and interesting problem concerning discrepancy hypotheses is the measurement of the discrepancy itself, that is, the distance between I and those optimally situated stimuli presumably most effective in mediating behavior.

Since there are too many free parameters to be initially estimated from data, it is suggested that distances among stimuli along the *J* scale be estimated by creating experimental conditions under which a Coombs function can be obtained. Once these parameters have been estimated, experimental conditions can be created under which it can be expected that discrepancy hypotheses will hold, and then the remaining parameters can be estimated. An empirical example is described below, but consider first the rationale for this suggested approach. What follows is based on Coombs' discussion of the unfolding model (Coombs, 1964, Ch. 5 and 9).

A Coombs function is illustrated in Figure 1C. In this preference model, stimuli which are coincident with I, if any, are most preferred, the remaining stimulus preferences correspond "to the rank order of absolute distances of the stimulus points from the ideal point [Coombs, 1964, p. 80]." Although this change from the discrepancy models seems slight, there are two important implications of this strategy. First, while the particular preference order relations expected under discrepancy models can vary with changes in the shape of the preference curve, in this model preferences are insensitive to changes in shape. Provided the curve remains symmetrical and continuously falls off about I, the same preference order relations will result. Second, while for both discrepancy models some preference order relations can imply different *J*-scale locations, here a preference order relation has a unique discrete region along the *J* scale. Under this model, preference order relations change only when a midpoint between two stimuli has been crossed (see Figure 1C—the dotted ordinates represent midpoints between stimulus pairs).

Viewed somewhat differently, the Coombs model is both a scaling method and a scaling criterion. Its use as a scaling criterion is seen as particularly important. As a criterion, the model permits an empirical test of the unidimensionality assumption. The primary criterion for unidimensionality (i.e., one-dimensional ordinal representation) is transitivity of the (preference) order relation (Torgerson, 1958, p. 247). As a scaling criterion, Coombs' model is an application of this test and requires transitive preferences. Given that preferences are transitive, a sufficient condition for rejection of a common unidimensional *J* scale is finding more than one mirror-image preference order relation (Coombs, 1964, p. 86). In Figure 1C, this preference order relation is ABC. Its mirror image is CBA. Note that data collected under conditions of the McClelland hypothesis as illustrated in Figure 1A reveal three mirror-image preference order relations. However, the Dember and Earl hypothesis, illustrated in Figure 1B, also reveals only one mirror preference order relation, the same as under Coombs' model. Viewed in this way, data collected to test hypotheses of the Dember and Earl model may be useful for answering the unidimensionality question.

As a scaling method, under some conditions, Coombs' model can extract metric information from the data. These procedures have been described by Coombs (1964, Ch. 5).

Once these *J*-scale stimulus distances have been determined, experimental conditions can be created under which a discrepancy model can be expected to hold, and ordinate heights to be used to determine the shape of the curve may be estimated from measures of strength of preference.

Consider the following example. In studies of exploration (Thomas, 1969b), cats wandered for several minutes in a triangular box with three compartments. A different visual pattern stimulus was in each compartment. It was hypothesized that responses to the set of stimuli would satisfy a unidimensional criterion. Time spent in each compartment was used to specify, for the cat, a preference order relation; the more time spent in a compartment, the more the pattern in that compartment was said to be preferred. One use of the preference order relations was to determine if the criterion for a unidimensional scale could be achieved. The hypothesis was supported.

No useful metric information can be extracted when three stimuli are used, so Coombs' model was used as a scaling criterion only. However, it was reasoned that by slightly adapting cats to selected stimuli, an I and stimulus point could be caused to coincide. By presenting *S*s with two choice stimuli, one on each side of I if Coombs' hypothesis holds, *S*s should prefer the closest stimulus. Using this procedure, distance order relations were obtained and a behaviorally defined metric was derived using Frank Goode's Δ solution (Coombs, 1964). To test the validity of the metric, additional cats were similarly tested, but predictions were made as to which of two stimuli should be preferred. The predictions were supported (Thomas, 1969a).

In additional research, cats have been adapted for longer periods of time with the same stimulus in all compartments. Then two choice stimuli were presented; the third compartment contained the same adaptation stimulus. Amount of time spent with the choice stimuli has been used to estimate the height of the stimulus ordinates. Knowing distances between the stimuli, discrepancies and parameters of the preference curve have been estimated. Strength of preference for different choice stimuli has been found to relate to length of adaptation—hardly a surprising finding. This suggests that discrepancy hypothesis preference curves are parameters, not constants.

DISCUSSION

The scaling approach suggested does have limitations, difficulties, and unstated assumptions, and it is important to consider some of them. An obvious practical difficulty is that the procedure is laborious. Yet there seems no obvious way to eschew the problem of parameter estimation since all parameters cannot be estimated jointly. Some conceptual simplification must take place which unfortunately increases empirical labor.

One working assumption is that the *J*-scale distances obtained under one experimental procedure will not, from the *S*'s viewpoint, "shrink" or "expand" when the experimental procedures are changed. This assumption may be false, but it can be tested experimentally. Another difficulty concerns Goode's Δ solution for numerically representing the ordered metric data. The procedure requires arbitrary decisions, and it is not clear if it is the best approach for producing metrics sturdy enough to withstand experimental verification. There are other numerical procedures available. The iterative-based procedures stemming from Shepard's (1962) pioneering efforts are obvious possibilities. However, these procedures were originally developed for similarities (QIV) data, not QI preference data, but Klemmer and Shrimpton (1963) have offered a modification of Shepard's procedure for preference data.

One reason for suggesting that stimulus distances be estimated under conditions favoring a Coombs function is that the Coombs model has no strong distributional assumptions. However, for some situations, the model may be inappropriate. It is an individual differences model which in the usual application involves assembling different preference order relations from different individual *S*s. The model would be useless if, for example, all *S*s gave the same preference order relation. A more likely situation, however, is that with many stimuli the preference order relations for different *S*s will be so varied that rejection of a common unidimensional *J* scale may be quite likely. Thus, the model is very vulnerable, that is, much data will probably not satisfy its conditions. Furthermore, it is not clear if the model is useful for grouped data.

Although there are conceptual differences between discrepancy models and the Coombs model, it is likely that under some conditions data could not distinguish the Coombs model from discrepancy models. This would depend on the distances separating stimuli, the I location, the depth of the "trough" as in Figure 1A, or the I′-to-pacer distance in Figure 1B, etc. As an example, stimuli to which *S*s were adapted which presumably fixed the *S*'s I location were not, generally, *most* preferred on choice trials as Coombs' model would predict (Thomas, 1969a); but this did not affect the hypotheses under test.

The paper has focused on the conceptually simple case of discrepancy models as unidimensional models. It has been argued that the question of unidimensionality should be a hypothesis to be answered in data and not an untested assumption. When the data are examined to see if unidimensionality is satisfied, the question being asked is whether the data satisfy the axioms of an ordinal scale. While transitivity of the order relation is not the only property of ordinal scales (cf., e.g., Pfanzagl, 1968, p. 29 and Ch. 4), it is probably the most important empirically testable property. However, it should be stressed that satisfying an ordinal criterion does not automatically reveal the stimulus attribute or attributes on which the psychological dimension is based. It may be that several different physical attributes are combined by the *S*s into one psychological dimension. Or it might be that no obvious physical attributes underlie a single psychological dimension. These are questions for experimental isolation. What is probably more likely is the data will not satisfy a unidimensional criterion. (In the writer's opinion, it is the probable failure of existing data to satisfy a unidimensional criterion that has caused the greatest single difficulty in evaluating from data the tenability of discrepancy hypotheses.)

There are at least two possible solutions to the problem if a unidimensional criterion cannot be achieved. One approach is to expand discrepancy models into the multidimensional case. This, however, leads to a number of difficult problems. For example, in the case of spatial models, one problem concerns the nature of the space. Should it be Euclidean or non-Euclidean? Furthermore, the problem quickly becomes conceptually complex as the number of dimensions increase beyond two or three. As far as is known, only Earl (1961) has considered the joint effects of two dimensions within a spatial–motivational framework. His formulation is necessarily tentative because experimental work on the problem is lacking.

A second alternative would be to reformulate discrepancy hypotheses within a choice theory framework (cf., e.g., Greeno, 1968). Some of the problems of spatial models would be eliminated, and choice theories may be able to handle more adequately than discrepancy theories psychologically interesting multiattribute stimuli such as symphonies, paintings, etc.

REFERENCES

BERLYNE, D. E. *Conflict, arousal, and curiosity.* New York: McGraw-Hill, 1960.

BLOCK, W. E. A preliminary study of achievement motive theory as a basis of patient expectations in psychotherapy. *Journal of Clinical Psychology,* 1964, **20**, 268–271.

CONNERS, C. K. Visual and verbal approach motives as a function of discrepancy from expectancy level. *Perceptual and Motor Skills,* 1964, 18, 457–464.

COOMBS, C. H. *A theory of data.* New York: Wiley, 1964.

DEMBER, W. N., & EARL, R. W. Analysis of exploratory, manipulatory, and curiosity behaviors. *Psychological Review,* 1957, 64, 91–96.

EARL, R. W. Problem solving and motor skill behaviors under conditions of free choice. Unpublished doctoral dissertation, University of Michigan, 1957.

EARL, R. W. *A theory of stimulus selection.* (Hughes Aircraft Company Ground Systems Special Document SD61-132) Fullerton, Calif.: Hughes Aircraft, 1961.

FELLER, W. *An introduction to probability theory and its applications.* Vol. 1. (3rd ed.) New York: Wiley, 1968.

FISKE, D. W., & MADDI, S. R. A conceptual framework. In D. W. Fiske & S. R. Maddi (Eds.), *Functions of varied experience.* Homeword, Ill.: Dorsey Press, 1961.

FOWLER, H. Satiation and curiosity. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation.* Vol. 1. New York: Academic Press, 1967.

GREENO, J. G. *Elementary theoretical psychology.* Reading, Mass.: Addison-Wesley, 1968.

HABER, R. N. Discrepancy from adaptation level as a source of affect. *Journal of Experimental Psychology,* 1958, **4**, 370–375.

HEBB, D. O. *The organization of behavior.* New York: Wiley, 1949.

HUNT, J. McV. Intrinsic motivation and its role in psychological development. *Nebraska Symposium on Motivation,* 1965, **14**, 189–282.

KAMMANN, R. Verbal complexity and preferences in poetry. *Journal of Verbal Learning and Verbal Behavior,* 1966, **5**, 536–540.

KLEMMER, E. T., & SHRIMPTON, N. W. Preference scaling via a modification of Shepard's proximity analysis method. *Human Factors,* 1963, **5**, 163–168.

McCALL, R. B., & KAGAN, J. Stimulus-schema discrepancy and attention in the infant. *Journal of Experimental Child Psychology,* 1967, **5**, 381–390.

McCLELLAND, D. C., & CLARK, R. A. Antecedent conditions for affective arousal. In D. C. McClelland, J. W. Atkinson, R. A. Clark, & E. L. Lowell (Eds.), *The achievement motive.* New York: Appleton-Century-Crofts, 1953.

PFANZAGL, J. *Theory of measurement.* New York: Wiley, 1968.

SACKETT, G. P. Effects of rearing conditions upon the behavior of rhesus monkeys (*Macaca mulatta*). *Child Development,* 1965, **36**, 855–868.

SCHRODER, H. M., DRIVER, M. J., & STREUFERT, S. *Human information processing.* New York: Holt, Rinehart & Winston, 1967.

SHEPARD, R. N. The analysis of proximities: Multidimensional scaling with an unknown distance function: I. *Psychometrika,* 1962, **27**, 125–140.

THOMAS, H. Preference following adaptation of an ideal and the metric scaling of visual complexity.

Journal of Experimental Psychology, 1969, **82**, 488–492. (a)

THOMAS, H. Unidirectional changes in preference for increasing visual complexity in the cat. *Journal of Comparative and Physiological Psychology*, 1969, **68**, 296–302. (b)

TORGERSON, W. S. *Theory and methods of scaling.* New York: Wiley, 1958.

TVERSKY, A., & RUSSO, J. E. Substitutability and similarity in binary choices. *Journal of Mathematical Psychology*, 1969, **6**, 1–12.

VERINIS, J. S., BRANDSMA, J. M., & COFER, C. N. Discrepancy from expectations in relation to affect and motivation: Tests of McClelland's hypothesis. *Journal of Personality and Social Psychology*, 1968, **9**, 47–58.

VITZ, P. C. Preference for different amounts of visual complexity. *Behavioral Science*, 1966, **11**, 105–114.

WALKER, E. L. Psychological complexity as a basis for a theory of motivation and choice. *Nebraska Symposium on Motivation*, 1964, **13**, 47–95.

WALKER, E. L., & WALKER, B. E. Response to stimulus complexity in the rat. *Psychological Record*, 1964, **14**, 489–497.

WOHLWILL, J. F. The physical environment: A problem for a psychology of stimulation. *Journal of Social Issues*, 1966, **22** (4), 29–38.

WOHLWILL, J. F. Amount of stimulus exploration and preference as differential functions of stimulus complexity. *Perception and Psychophysics*, 1968, **4**, 307–312.

APPENDIX

Let the integers 1, 2, 3, $\cdots$, n represent the N_a ordinally represented stimuli, and let k be the maximally preferred stimulus. If $k = 1$ or n there could be no single peaked preference function as defined above, for there could be no strictly monotonic falling off on both sides of k as required by definition. Thus to have a single peaked preference function, k must be some jth stimulus 2 through $n - 1$. To the left of j there are L stimuli; to the right, R stimuli; and preferences for these stimuli must fall off strictly monotonically from j to the left and right, respectively, as required by definition of single peaked preference function. This implies that there can be only one preference order for the L left stimuli and one preference order for the R right stimuli for any jth stimulus. It is clear that as j varies, L and R vary, and both L and R have limits of 1 and $N_a - 2$, and $L + R = N_a - 1$. These $N_a - 1$ stimuli may take any order for any jth stimulus provided the L left and R right stimuli retain their same order for that jth stimulus. The number of possible sequences of the $N_a - 1$ stimuli with this constraint is $\binom{L+R}{L}$ (cf. Feller, 1968, p. 36, Example e). Suppose $j = 2$; then $L = 1$, and $R = N_a - 2$. There are $\binom{N_a - 1}{1}$ possible orders for the $N_a - 1$ stimuli. For $j = 3$ there are $\binom{N_a - 1}{2}$ possible orders for the $N_a - 1$ stimuli. Generalizing and summing yields $\sum_j \binom{N_a - 1}{j - 1}$ which is algebraically equivalent to $2^{(N_a - 1)} - 2$, the numerator of Equation 1.

(Received June 22, 1970; revision
received December 14, 1970)

Psychological Review
1971, Vol. 78, No. 3, 260–261

THEORETICAL NOTES

COMMENTS ON BIEDERMAN'S "CONTINUITY THEORY REVISITED: A FAILURE IN A BASIC ASSUMPTION"

DANIEL B. BERCH[1]

California State College, Los Angeles

Biederman has argued that a basic assumption of Spence's discrimination learning theory is incorrect. It is suggested that one can account for Biederman's data on the basis of the stimulus generalization of habit while maintaining that the inhibition associated with the negative stimulus increases monotonically over the course of learning. It is concluded that Biederman's evidence is insufficient for rejecting Spence's theory.

Biederman (1970) has suggested that the validity of Spence's (1936) theory is suspect because it incorporates an incorrect assumption regarding the growth of habit strength in discrimination learning. Biederman's argument is based primarily on evidence from a series of experiments in which subjects were given concurrent but unequal amounts of training on two simultaneous discrimination problems, followed by a choice between the overtrained and nonovertrained negative stimuli. After relatively few training trials, the subjects showed a preference for the overtrained negative stimulus (Biederman, 1967, 1968; Deutsch & Biederman, 1965). According to Biederman, this finding suggests that in discrimination learning, the negative stimulus (S−) loses its aversiveness as training proceeds, that is, that the inhibition associated with the S− is nonmonotonic.

In the first place, Biederman's use of the term "overtraining" is inappropriate. Overtraining has conventionally referred to the administration of additional trials beyond attainment of some specified criterion. By this standard, most of the "more trained" problems in Biederman's experiments would not be regarded as overtrained; indeed, some cannot even be considered mastery trained.

Second, Biederman states that "the basic assumption" of Spence's (1936) theory which is incorrect is that habit strength increases monotonically with number of reinforcements. However, the bulk of Biederman's evidence, that which shows a preference for the more trained S−, has no bearing on this assumption, since responses to the negative stimulus are *never reinforced* in the simultaneous discrimination problem.

Third, although Biederman's evidence is relevant to Spence's assumption that the inhibition associated with the S− is a monotonic function of the number of nonreinforced trials, Biederman incorrectly translates this assumption to avoidance terms. Part of Biederman's confusion stems from his failure to recognize that Spence viewed inhibition as producing a decrement in the approach tendency rather than an increment in an avoidance tendency. A careful reading of Spence's (1936) article makes this distinction readily apparent.

Fourth, and most important, an alternative interpretation of Biederman's data is possible. Namely, the more trained S− may be preferred because it possesses a stronger *generalized* habit (approach) tendency than the less trained S−. This should occur as a result of the stronger habit tendency conditioned to the more trained positive stimulus (S+). Furthermore, assuming that inhibition is a monotonic function, the more trained S− should be preferred only when its generalized habit tendency is strong enough, relative to that of the less trained S−, to withstand the greater decrement due to its stronger inhibitory tendency. In other words, this line of reasoning suggests that the probability of choosing the more trained S− should vary as a function of factors which influence the relative strengths of the generalized habit tendencies associated with the two negative stimuli. An example of such a variable is the degree of similarity of the nonspatial cues within each of the discrimination problems. In any event, this analysis demonstrates that one can account for Biederman's results while maintaining that the inhibition associated with the S− increases montonically over the course of learning.

---

[1] Requests for reprints should be sent to Daniel B. Berch, Department of Psychology, California State College at Los Angeles, 5151 State College Drive, Los Angeles, California 90032.

It should be noted that Biederman (1968) has already rejected generalization of habit as an explanation of his results, because "the discriminative performance might be retarded [p. 462]." However, the fact that generalization of habit from S+ to S− would impede acquisition of the original discrimination clearly does not impugn the validity of this process as an explanation of the probe-trial findings.

It must be concluded that Biederman's evidence does not constitute a sufficient basis for rejecting Spence's (1936) theory. Inasmuch as Biederman's probe-trial technique may reflect in part the generalized habit tendency of the negative stimulus, this measure may prove to be inadequate for evaluating the monotonicity of inhibition per se.

REFERENCES

BIEDERMAN, G. B. The overlearning reversal effect: A function of the nonmonotonicity of the negativity of S− during discriminative training. *Psychonomic Science,* 1967, **7**, 385–386.

BIEDERMAN, G. B. Stimulus function in simultaneous discrimination. *Journal of the Experimental Analysis of Behavior,* 1968, **11**, 459–463.

BIEDERMAN, G. B. Continuity theory revisited: A failure in a basic assumption. *Psychological Review,* 1970, **77**, 255–256.

DEUTSCH, J. A., & BIEDERMAN, G. B. The monotonicity of the negative stimulus during learning. *Psychonomic Science,* 1965, **3**, 391–392.

SPENCE, K. W. The nature of discrimination learning in animals. *Psychological Review,* 1936, **43**, 427–449.

(Received July 15, 1970)

Psychological Review
1971, Vol. 78, No. 3, 262–271

HYPOTHESIS ANALYSIS OF CONJUNCTIVE CONCEPT-LEARNING SITUATIONS

LANCE A. MILLER[1]

IBM Thomas J. Watson Research Center, Yorktown Heights, New York

For a taxonomy of conjuctive concept-learning situations, mathematical expressions are presented giving the number of possible hypotheses concerning the problem rule as a function of the task informational variables. The number of initially possible hypotheses is shown to be a function of preproblem knowledge of the number of relevant dimensions and the type of initial focus stimulus, as well as the problem parameters. During problem solution, reduction in the number of possible hypotheses is shown to be a function of the interstimulus similarity. Optimal stimulus selection procedures and expressions for the expected trial of solution are derived for various task situations involving the selection paradigm.

The purpose of this paper is to present the analytical means for determining the number of possible concept rules, or hypotheses, in concept-learning situations as a function of the problem parameters, pretask information, and the stimulus sequences. This work is essentially a development and augmentation of the original contribution of Hovland (1952), who was one of the earliest proponents of an information-processing or "hypothesis search" point of view. However, in the present treatment, only conjunctive rules are considered, while Hovland was also concerned with a kind of disjunction rule (where any of several values of a particular dimension could characterize positives). This work subsumes both Hovland's concern for a hypothesis description of situations in which the subjects know in advance the number of relevant dimensions, and also that section in Bruner, Goodnow, and Austin's (1956) work investigating human concept learning, where mathematical considerations were given to a situation of prior ignorance of this information.

The following notations are employed throughout:

N_h = the number of logically possible (equiprobable) hypotheses concerning the correct conjunctive rule.
$U(N_h)$ = the amount of hypothesis uncertainty ($\log_2 N_h$, in bits).
n = the number of nonredundant stimulus dimensions.
v = the number of values representing a dimension (equal for all dimensions).
r = the number of relevant dimensions (dimensions for which positives have the the same value) in the experimental rule ($1 \leq r \leq n$).
d = in a comparison of two stimuli, the number of dimensions having different respective values ($o \leq d \leq n$).

N_h AT THE OUTSET OF A CONCEPT-LEARNING PROBLEM

Two types of information may be supplied to subjects in the pretask situation which determine the intital value of N_h, given that subjects know the rule is conjunctive. One type of information is that relating to problem complexity. Assuming that subjects are always informed of the number of stimulus dimensions (n) and the number of values per dimension (v), this type of information refers to whether subjects know in advance the number of relevant dimensions (r). When subjects know r in advance, there is Relevance Certainty (*RC*). This condition was assumed in Hovland's analyses. When r is unknown, there is Relevance Uncertainty (*RU*), and most experimental investigations have involved this condition. The second type of prior information is the presentation to the subject of an initial classified "focus" stimulus (cf. Bruner et al.). The three alternatives with respect to this variable are (*a*) positive focus, (*b*) negative focus, and (*c*) no focus (performance is measured from presentation of the first stimulus).

The three focus conditions, combined with the two conditions of *RC* and *RU*, yield six possible pretask conditions. Expressions for each are presented in Table 1, giving N_h as a

---

[1] The author thanks Robert C. Haygood, Michael I. Posner, and Robert W. Schvaneveldt for their helpful comments and criticisms on an early draft of this paper. Appreciation is also expressed to Robert C. Calfee and Jane Morantz for their helpful discussion concerning mathematical derivations.

Requests for reprints should be sent to Lance A. Miller, Thomas J. Watson Research Center, Yorktown Heights, New York 10598.

TABLE 1

FORMULAE GIVING PRETASK N_h AS A FUNCTION OF PROBLEM PARAMETERS FOR THE KNOWLEDGE OF r AND FOCUS CONDITIONS

Focus condition	Knowledge of r			
	RU		*RC*	
No focus	$N_h = (v+1)^n - 1$	[1]	$N_h = \binom{n}{r} v^r$	[2]
Positive focus	$N_h = 2^n - 1$	[3]	$N_h = \binom{n}{r}$	[4]
Negative focus	$N_h = (v+1)^n - 2^n$	[5]	$N_h = \binom{n}{r} (v^r - 1)$	[6]

function of the problem parameters n, r, and v. The derivations for these and the formulae to follow require only a straightforward application of combinatorial mathematics and are outlined in the Appendix.

With respect to the *RC–RU* variable, examination of the expressions makes possible the following predictions for human concept learning, assuming that the concept-learning processes involve an active search for the correct hypothesis, where the same hypothesis search strategy is employed for both conditions: (*a*) performance should be better for *RC* than for comparable *RU*; (*b*) manipulation of r should affect performance only for *RC*; (*c*) performance trials should be maximum for those r conditions of *RC* which maximize N_h; and (*d*) performance under *RC* and *RU* should be equivalent when the separate conditions are equated with respect to N_h by manipulation of the problem parameters. While *RC–RU* comparisons are of theoretical interest in view of these analyses, this variable has been employed in only one known study, where general support for these predictions was found, although this variable apparently exercised some control over the strategies selected by the subjects (Miller, 1969).

Examination of the formulae for the three focus conditions also reveals some interesting characteristics. Not only is the positive focus case most effective in reducing the set of possible hypotheses, but also only for this condition is N_h independent of v, and thus independent of the number of stimuli (when manipulated by variation of this parameter). The values of N_h for a negative focus and no focus depend primarily on and tend to converge with v. Of these two cases, however, the condition of no focus is of least practical interest, since either the positive focus or negative focus situation exists after presentation of the first classified stimulus. Similarly, the negative focus condition essentially resolves into that of the positive focus upon presentation of the first positive stimulus. However, the negative focus condition is still of theoretical interest for investigations concerned either with the utilization of negative information in problem solution (e.g., Cahill & Hovland, 1960), or the procedures adopted by subjects in the selection paradigm, where subjects select stimuli to learn of their classification prior to the first positive instance (e.g., Miller, 1969).

The two preproblem variables may also be directly compared in terms of the initial value of N_h resulting for various combinations of these variables. In general, a positive focus under *RU* results in a smaller value of N_h than for *RC* with no focus, and this difference increases with the n and v. The primary exception occurs for $r = 1$ (and $n > 3$), where the value of N_h for *RC* can be much smaller than for *RU* and a positive focus, and this difference increases with n and decreases with v. The minimum value of N_h, however, occurs for the most informative conditions of both of these variables—that is, positive focus and *RC*.

A final comment concerning the calculation of N_h at the problem outset relates to the possibility of redundant stimulus dimensions. While partially correlated dimensions do not influence the present expressions, completely correlated dimensions count only as a single independent dimension. For example, if the values of three dimensions (e.g., number, height, and color) are completely correlated, and one value combination characterizes positives (e.g., three tall red figures), these dimensions count as only one relevant dimension in the Table 1 formulae. The same holds true for completely correlated irrelevant dimensions.

N_h FOR DIFFERENT KINDS OF STIMULUS SEQUENCES

The concern of this section is the determination of N_h as stimulus information permits the elimination of hypotheses from the initial set, as given by the expressions in the preceding section. At the outset, it should be pointed out that there is no known exact analytical solution for determining the reduction of N_h for a randomly generated sequence of positives and negatives—the kind of sequence most frequently employed in experimental research. Nevertheless, analytical expressions having general implications are presented here for certain restricted kinds of sequences which are of interest in their own right.

Simultaneous Scanning Strategy

In the previous section, utilization of pretask information was assumed to be independent of the choice of a particular problem strategy. Here, however, it is desirable to determine the maximum possible reduction in N_h as a function of stimulus information, and the optimally efficient simultaneous scanning strategy identified by Bruner et al. is the appropriate reference choice.

The general operating procedure of this ideal strategy, as illustrated by an example in Table 2, is first to generate all the possible hypotheses, given the pretask information, and then to eliminate hypotheses from this set on the basis of information contained in subsequent stimuli. Positive stimuli are compared to each other, and hypotheses are eliminated which specify as relevant the noncommon dimensions among positives (cf. Trials 1 and 3 of the example). Negative stimuli are informative to the extent that they have values in common with positives. All combinations of such common values (for *RU*) or all *r*-tuples (for *RC*) are eliminated as hypotheses (cf. Trials 2, 5, and 6). This information processing is continued until there is only one remaining hypothesis.

This strategy may be contrasted to a "focusing" strategy (cf. Bruner et al.) in which hypotheses are never directly considered but information is gained by logical deductions concerning the relevance or irrelevance of dimensions. In the presently formulated strategy of *optimal focusing*, information from positives is utilized by logically intersecting these stimuli to determine the irrelevant dimensions. Negatives are compared to a working hypothesis which initially presumes the relevance of the values present in the first positive (e.g., "0000" for the Table 2 example), where these entries are appropriately modified as dimensions are deduced to be irrelevant or relevant (e.g., "$0\underline{0}0\text{x}$" for Trial 3 of the Table 2 example, where the underlining indicates a deduced relevant dimension). When a negative differs in only one value from the nonirrelevant entries in this working hypothesis, the relevance of the noncommon dimension may be deduced. The efficiency of this strategy (estimated in unpublished computer-simulation work to be within 6% of simultaneous scanning performance) derives both from the fact that negatives can be informative even when they differ in more than one value from positives (unlike other focusing strategies; cf. Bruner et al.) and from the capability of utilizing previously uninformative negatives when new irrelevant dimensions are determined—that is, previous negatives are reprocessed.

TABLE 2

RESULTS OF THE OPERATION OF SIMULTANEOUS SCANNING FOR A SAMPLE PROBLEM INVOLVING FOUR DIMENSIONS, *RC* ($r = 2$), AND POSITIVE FOCUS

Trial no.	Stimulus	Classification[a]	Eliminated hypotheses	Remaining possible hypotheses[b]
1	0 0 0 0	P	—	00xx, 0x0x, 0xx0, x00x, x0x0, xx00
2	0 1 0 0	N	0x0x, 0xx0, xx00	00xx, x00x, x0x0
3	0 0 0 1	P	0xx0, x0x0, xx00	00xx, x00x
4	1 1 0 1	N	—	00xx, x00x
5	0 1 0 1	N	0x0x	00xx, x00x
6	1 0 0 1	N	x00x	00xx

[a] P = positive; N = negative.

[b] The symbol "x" indicates irrelevant dimension in the hypothesis; the problem involves two relevant dimensions.

A compound of both these strategies is possible in which the simpler optimal focusing strategy is employed whenever possible, and simultaneous scanning is used only when relevance deductions cannot be made. The only assumption made in the following treatment is that the strategy employed utilizes all the available information from all stimuli to maximally reduce the uncertainty concerning the correct hypothesis.

N_h, *for an All Positive Sequence*

The most important parameter influencing the values of N_h for this and other sequence types is *d*, the number of dimensions for which two stimuli have different values. For positive stimuli, the range of *d* is from 1 to $(n - r)$, the number of irrelevant dimensions. Expressions 7 and 8 in Table 3 give the formulae for determining N_h in a two-stimulus sequence involving a positive focus and a subsequent positive of difference *d* from the focus. As indicated by Expression 8, when the second positive has $d = (n - r)$ differences from the focus in the *RC* condition, all but the correct hypothesis may be eliminated. When $d < (n - r)$, however, more than one positive following the focus will be required to permit deduction of the correct rule. In particular, when $d = 1$, and a new dimension is varied by each subsequent stimulus, a sequence of $(n - r)$ positives following the positive focus

TABLE 3

FORMULAE GIVING N_h AS A FUNCTION OF PROBLEM PARAMETERS FOLLOWING A POSITIVE FOCUS, FOR THE KNOWLEDGE OF r AND CLASSIFICATION OF THE SECOND STIMULUS CONDITIONS

Classification of second stimulus	Knowledge of r	
	RU	*RC*
Positive	$N_h = 2^{n-d} - 1$ [7]	$N_h = \binom{n-d}{r}$ [8]
Negative	$N_h = 2^n - 2^{n-d}$ [9]	$N_h = \binom{n}{r} - \binom{n-d}{r}$ [10]

is required for deduction of the correct hypothesis.

In contrast to the *RC* condition, with *RU*, a sequence of positives, in which all irrelevant dimensions are varied, results not in elimination of all but the correct hypothesis, but leads to reduction to some minimum set which cannot be reduced further by positive stimuli alone. This is shown by the following expression:

$$\text{Minimum } N_h \text{ (}RU\text{, positives)} = 2^r - 1 \quad [11]$$

N_h, *for a Positive Focus and Subsequent Negatives*

In a negative stimulus following a positive focus, at least one of the relevant dimensions must be different, and all dimensions may be different, such that $1 \leq d \leq n$. It can be seen from Expressions 9 and 10 in Table 3 that the largest number of hypotheses is eliminated when $d = 1$, while no hypotheses are eliminated when $d = n$ for *RU*, or when $(n - d)$ is less than r for *RC*. Unlike a positive sequence, the most informative negative following a positive focus does not lead to elimination of all but the correct hypothesis in *RC* (except when $r = 1$). A negative with $d = 1$ permits reduction of N_h to those hypotheses which specify as relevant the noncommon dimension between the negative and the focus (with the relevant dimensional value being that occurring in the focus). Complete reduction of N_h by means of negative stimuli following a positive focus, therefore, requires at least r negatives with $d = 1$, where a different noncommon dimension is varied by each.

In the *RU* condition, just as for a sequence of positive stimuli, a sequence of negatives following a positive focus reduces N_h only to some minimum value, which cannot be reduced by additional negative information:

$$\text{Minimum } N_h \text{ (}RU\text{, negatives following positive focus)} = 2^{n-r} \quad [12]$$

N_h, *for an All Negative Sequence*

For a sequence involving no positives, the informational value of a subsequent negative following a negative focus increases with d, rather than decreases as in the case of a positive focus and subsequent negatives. The initial negative focus eliminates $\binom{n}{r}$ or $2^n - 1$ hypotheses from the initial set given by Expression 2 (*RC*) or Expression 1 (*RU*), respectively. However, only in the *RC* condition for $r = 1$ or $r = n$ can subsequent negatives continue to eliminate this maximum number of hypotheses (a total of v such stimuli can be maximally informative for $r = 1$, while every negative is informative for $r = n$). For *RU*, or, generally, for $1 < r < n$, some smaller number of hypotheses is eliminated due to the fact that at some point in the sequence (depending primarily on v), subsequent negatives cannot be chosen without overlapping dimensional values of previous negatives, resulting in the redundant elimination of hypotheses (e.g., compare the negatives of Trials 2 and 5 in Table 2).

At present, there is no known general solution for determining the change in N_h as a function of n, r, and v for random negative sequences. Exact determination of the changing N_h values for a particular negative sequence (or a mixed sequence, for that matter) presently can be accomplished only by simulation of an optimal strategy for the sequence of interest and computing N_h on each trial.

Estimation of N_h *in a Mixed Sequence*

Despite the necessity of simulation for determining the values of N_h in a mixed sequence, some analytical assistance can be provided for estimating N_h in a positive focus sequence with following negative or mixed stimuli. The analyses require that the relatively simpler optimal focusing strategy be simulated up to and including the kth stimulus when N_h is to be estimated so that the number of deduced relevant and irrelevant dimensions can be determined. The author has found that this can usually be accomplished quickly and easily by "hand" simulations (rather than computer simulation) by beginning with an intersection of the positives on or preceding the trial for which N_h is to be estimated to compute the irrelevant entries for the working hypothesis. Previous negatives are then compared to the resulting working hypothesis to determine how many relevance deductions can be made (e.g., how many previous negatives differ uniquely from the working hypothesis in only one dimensional value).

Given that the simulation results in the knowledge that i dimensions are irrelevant and j dimensions are relevant, an estimate of the number of hypotheses remaining after k stimuli, $N_h{}^k$, for RC and RU is given by:

$N_h{}^k$ (RC, optimal focusing)

$$= \binom{n-i-j}{r-j} \quad [13]$$

$N_h{}^k$ (RU, optimal focusing)

$$= 2^{n-i-j} - f(j) \quad [14]$$

where $f(j)$ in Expression 14 is defined to be 1 if $j = 0$, and 0 if $j > 0$.

These formulae may occasionally underestimate N_h (particularly for small k), since account is not taken of potential information in negatives which differ in more than one value from the working hypothesis (whereas the simultaneous scanning strategy would utilize all information). The formulae given in Table 3 may also be modified to permit estimation of N_h on Trial $(k+1)$ as a function of d by substituting $(n-i-j)$ for n, and $(r-j)$ for r. It should be noted, however, that reduction of N_h by the next stimulus can occur only if the stimulus contains all j dimensional values deduced to be relevant, and if the stimulus is different in at least one of the values of the $(n-i-j)$ dimensions whose status is unknown.

Distribution of Stimulus Information as a Function of d

Expressions 7–10 indicate the general relationship that the informational value of successive stimuli is determined by their similarity (dimensional overlap) with previous stimuli. However, there may be several different stimuli at any point in a sequence which have the same potential informational value, and it is of interest to examine how the total amount of problem information is distributed over the stimulus population.

The following Expressions (15 and 16) give the frequency of second stimuli which have d differences from a positive focus, where the second stimulus is positive and negative, P_d and N_d, respectively (the superscript indicates the stimulus number):

$$P_d{}^2 = (v-1)^d \binom{n-r}{d} \quad [15]$$

$$N_d{}^2 = (v-1)^d \left[\binom{n}{d} - \binom{n-r}{d}\right] \quad [16]$$

Expression 15 indicates, for example, that for $n = 6$, $r = 3$, and $v = 2$, there are three possible second positives which have $d = 2$, any one of which could reduce the initial $U(N_h)$ by 2.3 bits of information in the RC condition. Similarly, solution of Expression 16 gives three alternative negatives having $d = 1$, each of which would result in a one-bit reduction in uncertainty.

Examination of the dependence of the frequency values on the value of v indicates that for second positives, intermediate informational values, $(n-r)/2 \leq d \leq (n-r-1)$, are most probable for small v, while for large v (say, $v > 8$), the most informative positives are also the most probable—that is, positives having $d = n - r$ are most frequent. For negatives, however, the least informative instances become the most probable as v increases.

Similar expressions may be given for the frequency of stimuli of a given class as a function of d on the kth trial, where i irrelevant and j relevant dimensions have been previously deduced (by the optimal focusing or similar strategy). The formulae given below permit determination of the number of stimuli on the kth trial which possess all of the j known relevant dimensional values and differ in one or more of the $(n-i-j)$ remaining dimensions whose status has yet to be determined. Stimuli not meeting these conditions are uninformative.

$$P_d{}^k = (v-1)^d \binom{n-r-i}{d} x\, v^i \quad [17]$$

$$N_d{}^k = (v-1)^d \left[\binom{n-i-j}{d} - \binom{n-r-i}{d}\right] x\, v^i \quad [18]$$

To illustrate how the distribution of information in the remaining stimulus set changes as the status of dimensions becomes known, consider an RC example where $n = 8$, $r = 4$, and $v = 2$. Following the positive focus there are initially four positive or negative stimuli with

$d = 1$ which convey one bit of information. However, at the point where the status of $i = j =$ two dimensions is known, there are eight $d = 1$ stimuli of both classes, each still conveying about one bit of information. In general, it appears that as the amount of uncertainty decreases in a problem, there is an increase in the probability that succeeding stimuli will be more informative, where this probability depends primarily on i for fixed v.

Optimal Selection Strategies

Consideration of the role of d in the above analyses leads to the identification of three procedures for selecting stimuli which are optimal for different task situations. Associated with two of these are formulae giving the expected trial of solution, assuming that all of the information available in the stimuli is utilized.

The first situation of concern is when there is only one relevant dimension, and this is known by the subjects. Solution of Expressions 8 and 10 for $r = 1$ indicates that maximum information results when $d = (n - 1)$, if the second stimulus is positive, and when $d = 1$, if the second stimulus is negative. Since the probabilities of both classes are equal (and assuming that the prior probabilities of dimensions being relevant are equal), the choice of d which maximizes reduction of N_h for either outcome is an average of the two solutions, resulting in $d = n/2$. Thus, in this situation, the best selection procedure, called simple optimal selection (since the concept is a simple unidimensional one), is to pick a second stimulus in which half of the dimensional values in the first stimulus are changed (either positive or negative focus), and successive stimuli vary half of the remaining untested dimensional values on each trial. In this way N_h is reduced by half by each stimulus, and solution will occur with $\log_2 n$ trials (this is the classic information-theory situation; cf. Garner, 1962). The extent to which subjects employ this selection strategy is not known, however.

The second selection strategy is appropriate for the conditions of RU and RC for $r > 1$. Assuming that the probability distribution for choosing a value of r is uniform over the range $1 < r < n$, negatives will, on the average, be much more probable than positives. One approach to the selection problem for this situation, then, is to determine the procedure which maximizes information gained, given the expectation of a negative stimulus. Examination of the Table 3 expressions indicates that a value of $d = 1$ maximizes the information in negatives (i.e., a relevant dimension may be deduced for a negative outcome) and leads to an equivalent amount of information if the outcome is positive (i.e., an irrelevant dimension may be deduced). This procedure, called optimal selection here, in connection with focusing utilization of information, corresponds to the conservative focusing strategy identified by Bruner et al. and has been observed in several other studies (e.g., Laughlin, 1965; Miller, 1969).

Assuming optimal selection (and complete information utilization), formulae can be derived giving the expected trial of solution, $\hat{T}$, for the RU and RC conditions (both with positive focus).

$$\hat{T}\ (RU,\ \text{optimal selection}) = U(N_h) = \log_2(2^n - 1) \approx n \qquad [19]$$

$$\hat{T}\ (RC,\ \text{optimal selection}) = r(n-r)(n+2)/(r+1)(n-r+1) \qquad [20]$$

The last situation of interest is that of a negative focus and RU. Here the problem has two possible formulations, depending on the stimulus sequence: (*a*) if the sequence can consist of only negative stimuli, how can stimuli be chosen experimentally to minimize $U(N_h)$ at each point in the sequence; and (*b*) if there are no constraints on the occurrence of positives, how can the subject select stimuli so as to minimize the number of trials before encountering a positive stimulus (at which point optimal selection is appropriate). The procedure discussed below is optimal for both formulations, but is considered with respect to the former.

The correct selection of the next negative stimulus, based on the previous consideration of negative sequences, is one which minimizes the overall similarity of the stimulus to each of the previous negatives. In general, successive sets of stimuli are to be chosen so that the number of common dimensions among members of the set and between stimuli of the previous set are minimized. The difficulty of applying this principle depends on n and v and is simplest (for $v > 2$) when $v = n$ (or at least when $n = 0$ modulo v). An example of an optimal sequence for $n = v = 3$ follows, where the dimensionality of the hypotheses eliminated is given in parentheses: "000, 111, 222" (all one-dimensional, and some two- and three-dimensional hypotheses are eliminated); "012, 120, 201, 021, 210, 102" (all one-dimensional hypotheses are redundantly eliminated;

the remaining two-dimensional hypotheses are eliminated, and additional three-dimensional hypotheses are eliminated); "001, 112, 220, 010, 121, 202, 100, 211, 022" (the one- and two-dimensional hypotheses are redundantly eliminated; new three-dimensional hypotheses are eliminated, leaving only nine such remaining). It is evident that the expected length of a stimulus sequence produced by this principle increases with *r*. However, the author was unable to derive an analytical expression giving the expected minimum number of trials (as a function of the problem parameters) for eliminating all *r*-dimensional hypotheses or for encountering a positive stimulus.

The above procedure, called optimal negative selection, is much easier to verbalize for problems in which $v = 2$: for any first stimulus, the second has the opposite values for all dimensions; the third differs from the second in terms of $n/2$ (randomly chosen) dimensional values; the fourth is the complement of the third; etc. This variant of the strategy has been previously suggested by Kochen (1960) and was followed only infrequently by subjects in unconstrained negative focus situations (Miller, 1969).

DISCUSSION

The importance of *d* in determining the informational value of stimuli, as shown by the present analyses, has several implications for the often encountered practice in the experimental literature of distinguishing among stimuli simply in terms of their assignment to the two classes of positive and negative instances (cf. references below). First, the present analyses do in fact demonstrate that the expected informational value of randomly chosen positives is greater than that for negatives. Nevertheless, depending on *d* (and *v*), a particular set of positives may be either highly redundant or else permit maximum reduction in uncertainty (only two are necessary). Similarly, a set of negatives may either convey no information at all, or else lead to maximum uncertainty reduction with only *r* negatives following a positive focus.

Second, description of experimental conditions in terms of the proportion of positives in a sequence (e.g., Haygood & Stevenson, 1967; Schvaneveldt, 1966; Whitman & Garner, 1963) provides only partial definition of their informational characteristics, since sequences having a fixed proportion of positives may differ greatly in the manner in which the uncertainty changes as a function of trials. Furthermore, even sequences having similar such functions may differ in the opportunities available for employing different information-utilization strategies. When negatives always differ from positives (or the working hypothesis) in more than one dimensional value, only a hypothesis-testing strategy such as simultaneous scanning is able to utilize any information they convey. Similarly, sequences may determine the opportunity for utilization of different focusing strategies. Thus, when negatives are always different from positives in more than one dimension, only optimal focusing can be employed when the negatives differ in a single value from the working hypothesis. A less efficient focusing strategy which requires the single difference between negatives and positives cannot be used. It may be that the common finding that subjects do not utilize negative information efficiently (cf. Bourne, 1966) can be attributed to the relative ease of a focusing procedure for using positive information as contrasted to a hypothesis-testing procedure possibly necessary to utilize the negative information, rather than to any marked inefficiency in utilizing negative information per se (cf. Davidson, 1969). Further insight into this particular problem will require comparison of performance for which the opportunity of utilizing various strategies is controlled.

Third, a similar difficulty is found with descriptions of task conditions in terms of the "availability" of previous instances (e.g., Bourne, Goldstein, & Link, 1964; Pishkin & Wolfgang, 1965). The availability of a single positive stimulus may always be expected to be of some use in that the set of possible common values among positives is thereby displayed. While the additional availability of other positives with $d > 0$ always guarantees the possibility that irrelevant dimensions may be deduced, mere availability of negative stimuli does not necessarily offer the possibility of corresponding relevance deductions, nor even the possibility that any information is thereby made available at all (e.g., given a positive focus "0000," and $v = 4$, the negatives "1123, 1231, 2312," etc.—81 in all—convey no information). In comparing the effects of availability of previous positives and negatives on performance efficiency, some account must be taken of the information made available by these stimuli and of the opportunity for utilizing different strategies given this information.

A topic related to the above question of positives versus negatives concerns the interstimulus similarity relationships of a stimulus sequence. The present analysis indicates that the optimal manipulation of the similarity between stimulus pairs is to make positive–positive (or negative–negative) pairs minimally similar, and positive–negative pairs maximally similar. Empirical support for this prediction is given by studies in which transfer performance was facilitated by the optimal type of variation as compared to less optimal conditions (Anderson & Guthrie, 1966; Detambel & Stolurow, 1956). A similar question, yet to be empirically tested, is whether the maximally informative sequences generated by the three optimal selection procedures facilitate performance as compared to less optimal (e.g., random) presentations.

In summary, the present analyses afford the opportunity for further characterizing conjunctive concept-learning tasks in terms of the information available to and the performance expected from an optimal problem solver. Analyses of this kind provide "absolute" intraexperimental and interexperimental standards for evaluating the efficiency of human performance. In the present case, the analytical expressions permit a uniform description of most concept-learning task situations in terms of the number of initially possible hypotheses. The expressions themselves detail the information potential of positive and negative stimuli as a function of both the initial problem conditions and the similarity among stimuli within a stimulus sequence. The present discussion of the experimental literature with respect to the analyses illustrates the necessity of providing highly specific informational descriptions of tasks as a basis for the further detailed understanding of the information-processing aspects of human performance. Finally, the expressions provide a means for generating a variety of informationally configured problem conditions for the testing of hypotheses concerning human concept learning.

REFERENCES

ANDERSON, R. C., & GUTHRIE, J. T. Effects of some sequential manipulations of relevant and irrelevant stimulus dimensions on concept learning. *Journal of Experimental Psychology*, 1966, **72**, 501–504.

BOURNE, L. E., JR. *Human conceptual behavior.* Boston: Allyn & Bacon, 1966.

BOURNE, L. E., JR., GOLDSTEIN, S., & LINK, W. E. Concept learning as a function of availability of previously presented information. *Journal of Experimental Psychology*, 1964, **67**, 439–448.

BRUNER, J. S., GOODNOW, J. J., & AUSTIN, G. A. *A study of thinking.* New York: Wiley, 1956.

CAHILL, H. E., & HOVLAND, C. I. The role of memory in the acquisition of concepts. *Journal of Experimental Psychology*, 1960, **59**, 137–144.

DAVIDSON, M. Positive versus negative instances in concept identification problems matched for logical complexity of solution procedures. *Journal of Experimental Psychology*, 1969, **80**, 369–373.

DETAMBEL, M. H., & STOLUROW, L. M. Stimulus sequence and concept learning. *Journal of Experimental Psychology*, 1956, **51**, 34–40.

GARNER, W. R. *Uncertainty and structure as psychological concepts.* New York: Wiley, 1962.

HAYGOOD, R. C., & STEVENSON, M. Effects of proportion of positive instances upon concept learning. *Psychological Reports*, 1967, **20**, 179–182.

HOVLAND, C. I. "Communication analysis" of concept learning. *Psychological Review*, 1952, **59**, 461–472.

KOCHEN, M. Experimental study of "hypothesis-formation" by computer. *IBM Research Report* (IBM Watson Research Center), 1960, RC-294.

LAUGHLIN, P. R. Selection strategies in concept attainment as a function of number of persons and stimulus display. *Journal of Experimental Psychology*, 1965, **70**, 323–327.

MILLER, L. A. *The effects of procedural and informational variables on conjunctive concept attainment performance.* (Doctoral dissertation, University of Wisconsin) Ann Arbor, Mich.: University Microfilms, 1969. No. 69-962.

PISHKIN, V., & WOLFGANG, A. Number and type of available instances in concept learning. *Journal of Experimental Psychology*, 1965, **69**, 5–8.

SCHVANEVELDT, R. W. Concept identification as a function of probability of positive instances and number of relevant dimensions. *Journal of Experimental Psychology*, 1966, **72**, 649–654.

WHITMAN, J. R., & GARNER, W. R. Concept learning as a form of internal structure. *Journal of Verbal Learning and Verbal Behavior*, 1963, **2**, 195–202.

APPENDIX

MATHEMATICAL DERIVATIONS

The derivations of the expressions, identified by the text numbering, are indicated very briefly below without reproduction of the final form. Comments given in parentheses refer to related previous work. The following mathematical identity is employed and referred to by letter.

$$\sum_{i=1}^{n} \binom{n}{i} = 2^n - 1 \qquad [a]$$

1. For each of the n dimensions which may be relevant, there are v possible relevant values.

Since a dimension may be irrelevant also, the total number of possible states for each dimension is $v+1$, and the nonpermissible case where all dimensions are irrelevant is subtracted.

2. There are $\binom{n}{r}$ ways of selecting r dimensions as relevant out of n variable dimensions; for each combination the r dimensions may be assigned any of v values. (This corresponds to Hovland's Expression 4.)

3. The correct concept is one of the combinations of i relevant values ($1 \leq i \leq n$) out of n possible values, or $\binom{n}{i}$, which is summed using Identity a. It is also derived by letting $v = 1$ in Expression 1. (Bruner et al. presented this expression in the nonsimplified summation form.)

4. The correct concept is one of the combinations of r values out of the n values present in the positive focus. It is also derived by letting $v = 1$ in Expression 2.

5. The negative eliminates $2^n - 1$ hypotheses from the set of $(v+1)^n - 1$ initially possible hypotheses.

6. The negative eliminates $\binom{n}{r}$ hypotheses from the set of $\binom{n}{r}v^r$ initially possible hypotheses.

7. The remaining set of hypotheses following a second positive is derived from all combinations of the $(n-d)$ values common to the two stimuli; summed using Identity a. (Bruner et al., p. 114, give an erroneous expression for the number of hypotheses eliminated.)

8. A second positive with d differences restricts the possible hypotheses to some combination of r out of the $(n-d)$ remaining common dimensions.

9. The negative permits elimination of hypotheses given by Expression 7 from the set given by Expression 3. (Bruner et al., p. 114, are slightly in error, as $2^{n-d} - 1$ hypotheses are eliminated rather than 2^{n-d}.)

10. The negative permits elimination of hypotheses given by Expression 8 from the set given by Expression 4.

11. The remaining hypotheses are all combinations of i out of the remaining r common dimensions, $1 \leq i \leq r$; summed using Identity a.

12. The minimum set is composed of those ombinations of the r relevant dimensions with he $i = 0, 1, \cdots, (n-r)$ remaining irrelevant dimensions; summed using Identity a, but from $i = 0$.

13. Each new deduction reduces the total number of dimensions out of which some combination may be selected. When only i irrelevant dimensions are deduced, there are $\binom{n-i}{r}$ remaining hypotheses; when only j relevant dimensions are deduced, there are $\binom{n-j}{r-j}$ remaining hypotheses. Both conditions are represented in the single expression.

14. The set of possible hypotheses is the combination of the remaining $(n-i-j)$ untested values taken $k = 0, 1, \cdot\cdot\cdot, (n-i-j)$ at a time; summed using Identity a but from $k = 0$. When no relevant dimensions are known, $f(j)$ is 1 since the hypothesis of all irrelevant dimensions is disallowed.

15 and 16. The total number of stimuli, v^n, may be expressed as a similarity function giving the number of stimuli of different d from any focus stimulus. For any d there are $\binom{n}{d}$ ways of choosing a stimulus with these differences, and for each combination, each of the d values may be chosen in $(v-1)$ ways. Thus, $v^n = \sum_{d=0}^{n} \binom{n}{d} (v-1)^d$. The combinatorial expression $\binom{n}{d}$ may be divided into two subexpressions, one for each of the stimulus classes. For positive stimuli there are $\binom{n-r}{d}$ ways of picking a second positive with d differences limited to a maximum of $(n-r)$ irrelevant dimensions—that is, positives must have r dimensions in common. For negative stimuli, out of the total of $\binom{n}{d}$ possibilities, the d differences may not be selected from just the $(n-r)$ irrelevant dimensions, or else the stimulus would be positive. These $\binom{n-r}{d}$ possibilities must, therefore, be subtracted from the total, and the negative component or $\binom{n}{d}$ is thus $\binom{n}{d} - \binom{n-r}{d}$.

17 and 18. For positives, there are only a total $(n-r-i)$ untested irrelevant dimensions from which the d differences may be selected. However, for each of the $\binom{n-r-i}{d}$ possibilities, the values of the irrelevant di-

mensions may take on any value, for a total of v^i ways. For negative stimuli, the d differences may be selected from any of the $(n - i - j)$ untested dimensions, but selection from just the $(n - r - i)$ untested irrelevant dimensions are not allowable, else the stimulus would be positive.

19. In the *RU* condition the testing of each successive dimension halves N_h on each trial such that $\hat{T} = n$, for the range $0 \leq r \leq n$. When, as in most problems, at least one dimension must be relevant, this expression is slightly inflated due to the possibility that, when $r = 1$, the problem can be solved in $(n - 1)$ trials if all of these lead to irrelevance deductions. Similarly, there is additional inflation when a maximum value of r—$r_{\max}$—is established and made known to the subjects. They can cease testing if they happen to discover all $r_{\max}$ relevant dimensions before the nth trial. For situations involving a value of $r_{\max}$ which is small compared to n (say $r_{\max} < n/2$), the exact waiting-time expression should be derived. This would consist of three components, one for each of the following situations: (*a*) the case where $r = 1$ (i.e., $(n - 1/n)/Pr(r = 1)$); (*b*) the case where $1 \leq r \leq r_{\max}$, where n is the proper estimate of $\hat{T}$ (weighted by $Pr\ (1 \leq r < r_{\max})$); and (*c*) the case where $r = r_{\max}$, where a component must be included to account for the possibility of solving in from $r_{\max}$ to $(n - 1)$ trials (cf. the left-hand component in the $\hat{T}$ expression given below).

20. In the *RC* condition, optimal selection may cease when either the r relevant dimensions or the $(n - r)$ irrelevant dimensions have been deduced. The waiting-time expression for $\hat{T}$ thus has two components. Solution may occur on the jth trial only if no more (and no less) than $(r - 1)$ relevant or $(n - r - 1)$ irrelevant dimensions were deduced on Trial $(j - 1)$ (where the remaining dimension is found on Trial j). This can occur in $\binom{j-1}{r-1}$ ways for the first condition and in $\binom{j-1}{n-r-1}$ ways for the second. The probability is obtained by dividing these combinatorial expressions by $\binom{n}{r}$, which gives the total number of ways r relevant dimensions may be chosen from n dimensions. The probability of solution on Trial j, times the value of j, then gives the trial weight for that situation, where summation is taken from r and $(n - r)$ for the two conditions to $(n - 1)$:

$$\hat{T} = \left[\sum_{j=r}^{n-1} \binom{j-1}{r-1} j + \sum_{j=n-r}^{n-1} \binom{j-1}{n-r-1} j \right] \Big/ \binom{n}{r}$$

The summation expressions may be eliminated by using the identity:

$$\sum_{j=r}^{n-1} \binom{j-1}{r-1} j = r \binom{n}{r+1}$$

and the final expression is obtained by expanding the combinational expressions and simplifying algebraically.

(Received July 17, 1970; revision received February 5, 1971)

Psychological Review
1971, Vol. 78, No. 3, 272–273

COMMENT ON "GENERAL SHAPE FUNCTION MODEL OF LEARNING WITH APPLICATIONS IN PSYCHOBIOLOGY"

JOEL P. WIESEN[1]

King's College, Wilkes-Barre, Pennsylvania

A reader of a recent study by Warburton and Greeno is likely to form a gross misconception of the probability of forming very dissimilar groups when the randomization process is applied with small groups. This is due to a misapplication of elementary probability theory. The model they presented was, in part, justified by this erroneous approach. The other theoretical considerations motivating their analysis are not affected by this error.

Warburton and Greeno (1970) present an innovative model and an involved technique for analyzing data from some standard learning paradigms. These are presented as an alternative to relying on the frequently unequal groups which are allegedly the unavoidable outcome when one forms small groups by the random assignment of subjects (*S*s) to groups. They state that this (alleged) weakness leads to the commission of Type I errors. Their analysis is also presented as an approach to some theoretical problems, such as distinguishing between learning curves of different shapes. This paper questions only the statements made by Warburton and Greeno about the limits of the randomization process.

Their lack of faith in the equality of groups in a randomized groups design is supported by some rather shocking probabilities. They state that when dealing with small groups (a common situation), the random assignment of *S*s to two groups may too often yield groups which differ greatly before the treatment is applied. This initial difference, they continue, is often not brought to light, since *S*s are often tested only after completing the experiment (i.e., after the independent variable is applied). If it were common to form such obviously biased groups, it would indeed be wise to consider randomization only for larger groups, and perhaps to look for a method (e.g., a model) which would enable one to cope with these differences when small groups are unavoidable. (Note, however, that the statistical procedures which require randomization take possible biases into account; Myers, 1966, p. 7.)

More specifically, they present the following quite plausible situation: An *E* has decided to form two groups of eight *S*s. It is assumed that of the 16 available, 8 are "bright" and 8 are "dull." To show how fallible randomization can be, they present the probability of one of the chosen groups consisting of eight *S*s of one kind, and the probability of one of the chosen groups having at least seven of one kind. (If one group formed from the 16 *S*s has at least 7 dull *S*s, the group consisting of the remaining *S*s must contain at least 7 bright *S*s. These are obviously two quite different groups.) The probability figures they give are 1/128 and 1/14, respectively.

How were these figures obtained? It seems that Warburton and Greeno assumed that the probability of choosing a bright *S* [$P(B)$] and the probability of choosing a dull *S* [$P(D)$] both equal $\frac{1}{2}$ for an infinite population of *S*s. Then the probability of choosing a group of eight of one kind from the total population would be $2/256 = 1/128$. Using the binomial expansion (Adler & Roessler, 1964), the probability of choosing a group of eight *S*s with at least seven alike would be $1/256 + 8/256 + 8/256 + 1/256$ (i.e., the probability of choosing a group of either eight bright, seven bright and one dull, seven dull and one bright, or eight dull) $= .0703 \simeq 1/14$. Thus, Warburton and Greeno seem to confuse their explicit example with an application of the randomization procedure to selection of *one* group from an infinite population with $\frac{1}{2}$ bright and $\frac{1}{2}$ dull animals. The following is their statement of the problem.

> One problem of learning experiments has been individual differences . . . the experimenter must attempt to remove any bias by the procedure of randomization. . . . However, with the small samples generally used in psychobiological research, randomization can result in arrangements systematic with respect to the uncontrolled factor. Suppose that in a group of 16 animals, half are "maze bright" and half are "maze dull," then there is a probability of 1 in 128 that a group of eight will be all of one kind, but only a probability of 1 in 14 that there will

---

[1] Requests for reprints should be sent to Joel P. Wiesen, Department of Psychology, King's College, Wilkes-Barre, Pennsylvania 18702.

TABLE 1

SEVERAL APPROACHES TO THE CALCULATION OF THE PROBABILITY OF FORMING BIASED GROUPS BY RANDOMIZATION

Approach	Type of groups to be formed	
	Eight bright in one group Eight dull in one group	At least seven bright in one group and at least seven dull in one group
Warburton & Greeno, 1970	$C_{8,0}P^8 + C_{8,8}Q^{8}$[a] 1/128[b] .00781	$C_{8,0}P^8 + C_{7,1}P^7Q + C_{8,7}PQ^7 + C_{8,8}Q^{8}$[a] 1/14[b] (actually 9/128)[a] .07143 (actually .07031)[a]
Infinite population of half bright and half dull *S*s	$(C_{8,0}P^8 \times C_{8,8}Q^8) + (C_{8,8}Q^8 \times C_{8,0}P^8)$ 1/32,768 .00003	$(C_{8,0}P^8 \times C_{8,8}Q^8) + (C_{8,8}Q^8 \times C_{8,0}P^8) + (C_{8,0}P^8 \times C_{8,7}PQ^7) + (C_{8,7}PQ^7 \times C_{8,0}P^8) + (C_{8,1}P^7Q \times C_{8,8}Q^8) + (C_{8,8}Q^8 \times C_{8,1}P^7Q) + (C_{8,1}P^7Q \times C_{8,7}PQ^7) + (C_{8,7}PQ^7 \times C_{8,1}P^7Q)$ 1/400 (actually 81/32,768) .00247
Finite population of 16 *S*s: 8 bright and 8 dull	$\frac{P_{8,8}}{P_{16,8}} \times 2$ 1/6,435 .00016	$\frac{P_{8,7} \times 9}{{}_{16}P_8} \times 2$ 1/715 .00140

[a] Inferred from Warburton and Greeno (1970).
[b] Actual values given in Warburton and Greeno (1970).

> be seven or more of one kind. In the usual experiment, of course, the information about the uncontrolled factor is not available, and so extreme random arrangements of small populations cannot be avoided [p. 348].

This states that given a finite population consisting of eight bright and eight dull *S*s, the probability is 1 in 14 that two groups will be chosen with at least seven bright in one and seven dull in the other. (Note again that if there are at least seven of one kind in one group, there must be at least seven of one kind in the other group.) Presuming random selection, these probabilities are grossly incorrect.

We can assess the probabilities of forming such groups starting with two different assumptions. First, we may randomly select from an infinite population of which half are bright and half are dull animals. Second, we may choose randomly from a finite population of 16 animals of which 8 are bright and 8 are dull. Warburton and Greeno seem to present the latter case. The probabilities associated with the chance formation of one group of eight bright *S*s and one group of eight dull *S*s are .00003 and .00016 for the two approaches, respectively. The probabilities associated with the chance formation of one group with at least seven bright *S*s and one group with at least seven dull *S*s are .00247 and .00140 for the two approaches, respectively. These probability values are presented in Table 1 (with their unsimplified solutions) along with the values given by Warburton and Greeno (with their probable origins). The probability of forming two such groups is thus not about 7% but, at best, $\frac{1}{4}$ of 1%.

It can be seen that the reason Warburton and Greeno could demonstrate a lack of theoretical support for the randomization procedure was due to a quite basic misinterpretation of a tricky problem and the ensuing incorrect probabilities.

REFERENCES

ADLER, H. A., & ROESSLER, E. B. *Introduction to problems and statistics.* (3rd ed.) San Francisco: Freeman, 1964.

MYERS, J. L. *Fundamentals of experimental design.* Boston: Allyn & Bacon, 1966.

WARBURTON, D. M., & GREENO, J. G. General shape function model of learning with applications in psychobiology. *Psychological Review,* 1970, **77**, 348–352.

(Received August 21, 1970)

Psychological Review
1971, Vol. 78, No. 3, 274

LETTERS TO THE EDITOR

RECOGNITION MEMORY AND SIGNAL DETECTION ANALYSES

Though it has been several years since Egan (1958) first applied the theory of signal detection (TSD) to recognition memory, the last few years have experienced a boom in such applications (see Banks, 1970). The most important contribution this has made has been to give specific recognition to the decision properties of recognition memory and to the role of subject-established criteria. Since recognition memory is so very prevalent in "real life" situations involving memory, this would appear to be an important contribution.

Tulving and Madigan (1970) have claimed the application of TSD to recognition memory to be not only unimportant but actually frivolous. They state that TSD "provides good research opportunities for those students who like exotic quantitive techniques . . . [p. 468]," and then go on to dismiss the application of TSD on the basis of a totally incorrect assertion, that "to the extent that the slopes of ROC curves depart from unity, measures of sensitivity and criterion are not independent [p. 468]." The various indexes of sensitivity discussed by Green and Swets (1966, pp. 96–98) for ROC curves whose slopes are not unitary are *all* criterion free. The independence of sensitivity and criterion measures is a fundamental tenet of TSD: the only way to obtain a lack of independence is to make an egregious error in determining the measure of sensitivity, such as looking up d' in Elliott's (1964) tables, which are based on the unitary slope assumption, when the actual slope of the ROC curve departs from unity.

Tulving and Madigan would prefer simple information about hits and false alarms, but comparisons of performance across tasks requires some correction for false alarms or criterion. The correction for false alarms suggested by Woodworth and Schlosberg (1954, p. 700) has been shown by Egan (1958) to be implicitly based on theoretical considerations that are greatly at odds with data. On the other hand TSD provides a *criterion-free* measure of sensitivity or memory.

Since the attack by Tulving and Madigan is based at least in part on a patently incorrect assertion about TSD, it seems necessary to set the record straight. Perhaps by the word "exotic," Tulving and Madigan mean merely something that they do not understand.

REFERENCES

Banks, W. P. Signal detection theory and human memory. *Psychological Bulletin*, 1970, **72**, 81–99.

Egan, J. P. *Recognition memory and the operating characteristic.* (Tech. Note AFCRC-TN-58-51) Bloomington: Indiana University, Hearing and Communication Laboratory, 1958.

Elliott, P. B. Tables of d'. In J. A. Swets (Ed.), *Signal detection and recognition by human observers.* New York: Wiley, 1964.

Green, D. M., & Swets, J. A. *Signal detection theory and psychophysics.* New York: Wiley, 1966.

Tulving, E., & Madigan, S. A. Memory and verbal learning. *Annual Review of Psychology*, 1970, **21**, 437–484.

Woodworth, R. S., & Schlosberg, H. *Experimental psychology.* New York: Holt, 1954.

(Received January 26, 1971)

Harley A. Bernbach
Department of Psychology
Cornell University
Ithaca, New York 14850

The British Journal of Psychology

Volume 62, Part 1 February 1971

CONTENTS

P. E. BRYANT Discrimination learning and the study of transfer in young children

B. M. HERMELIN AND N. O'CONNOR Children's judgements of duration

ANN L. BROWN AND BARBARA B. LLOYD Criteria of success in a developmental study of oddity learning

ROBERT RICH Constant ratio rule for confusion matrices from short-term memory experiments

ALEX MCGLAUGHLIN AND H. C. A. DALE Stimulus similarity and transfer in long-term paired-associate learning

MUKUL K. DEY Reminiscence in paired-associate learning

HOWARD R. POLLIO AND RUSSELL FOOTE Memory as a reconstructive process

P. D. MCLEOD, C. E. WILLIAMS AND D. E. BROADBENT Free recall with assistance from one and from two retrieval cues

CAROLE H. ERNEST AND ALLAN PAIVIO Imagery and sex differences in incidental recall

JANET F. FLETCHER AND AUBREY J. YATES The repetition of single speech sounds under delayed auditory feedback

R. A. M. GREGSON AND M. B. SIMMONDS Qualitative gustatory characteristics of disodium-5′-guanylate

J. P. DAS AND A. C. BOWER Orienting responses of mentally retarded and normal subjects to word-signals

JOHN T. HAWORTH Conditioned emotional response phenomena and brain stimulation

P. ROBERTOUX, M. CARLIER AND J. CHAGUIBOFF Preference for non-objective art: personal and psychosocial determiners

EDGAR MILLER Handedness and the pattern of human ability

NATHAN KOGAN A clarification of Cropley and Maslany's analysis of the Wallach-Kogan creativity tests

BOOK REVIEWS

£2.50 (US $8.50) each, plus postage. Annual subscription £7.50 (US $25.00)

CAMBRIDGE UNIVERSITY PRESS
200 EUSTON ROAD, LONDON NW1 2DB

9 NOV 1971

NOW AVAILABLE

A SOURCE BOOK FOR THE FIRST COURSE IN PSYCHOLOGY

A source book for teachers of general psychology, containing reviews of introductory texts, books of readings, and lab manuals; descriptions of psychological journals; a catalogue of audiovisual materials; a listing of psychological equipment manufacturers; bibliographies of collateral readings for students and reference works for teachers; names and addresses of national organizations; a discussion of teaching methods; and suggestions of ways to organize a first course in psychology. The book is one of several products of a project held at Oberlin College in the summer of 1970. Although the title of the book — *PROGRAM ON THE TEACHING OF PSYCHOLOGY IN THE SECONDARY SCHOOL: Final Report* indicates an applicability to a particular academic level, the book is useful in teaching general psychology regardless of the level.

Pages: 128 **Price: $2.00**

SOURCE BOOK ORDER FORM

Name ....

Address ....

....ZIP CODE...........

Send this form to:

American Psychological Association
Order Department
1200 - Seventeenth Street, N.W.
Washington, D. C. 20036

Orders must be prepaid. Single copy $2.00. Quantity discounts: 3-10 copies 10%, 11-Over copies 20%. Make checks payable to the American Psychological Association.

MADE BY THE DISABLED
REHABILITATION - INDIA
CALCUTTA—700 053